

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Barenkamp, Stephen J
- (ii) TITLE OF INVENTION: High Molecular Weight Surface Proteins of Non-Typeable Haemophilus
- (iii) NUMBER OF SEQUENCES: 11
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Shoemaker and Mattare, Ltd.
 - (B) STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza Bldg. 1
 - (C) CITY: Arlington
 - (D) STATE: Virginia
 - (E) COUNTRY: U.S.A.
 - (F) ZIP: 22202-0286
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/617,697
 - (B) FILING DATE: 01-APR-1996
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/302,832
 - (B) FILING DATE: 05-OCT-1994
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US PCT/US93/02166
 - (B) FILING DATE: 16-MAR-1993
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Berkstresser, Jerry W
 - (B) REGISTRATION NUMBER: 22,651
 - (C) REFERENCE/DOCKET NUMBER: 1038-557
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (703) 415-0810
 - (B) TELEFAX: (703) 415-0813

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 5116 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ACAGCGTTCT CTTAATACTA GTACAAACCC ACAATAAAAT ATGACAAACA ACAATTACAA	60
CACCTTTTTT GCAGCTATA TGCAAATATT TTA AAAAATA GTATAAATCC GCCATATAAA	120
ATGGTATAAT CTTTCATCTT TCATCTTTCA TCTTTCATCT TTCATCTTTC ATCTTTCATC	180

TTTCATCTTT CATCTTTCAT CTTTCATCTT TCATCTTTCA TCTTTCATCT TTCATCTTTT	240
ACATGCCCTG ATGAACCCAG GGAAGGGAGG GAGGGGCAAG AATGAAGAGG GAGCTGAACG	300
AACGCAATG ATAAAGTAAT TTAATTGTTC AACTAACCTT AGGAGAAAAT ATGAACAAGC	360
TATATCGTCT CAAATTCAGC AACGCCTGA ATGCTTTGGT TGCTGTGTCT GAATTGGCAC	420
GGGGTTGTGA CCATTCCACA GAAAAAGGCA GCGAAAAACC TGCTCGCATG AAAGTGCGTC	480
ACTTAGCGTT AAAGCCACTT TCCGCTATGT TACTATCTTT AGGTGTAACA TCTATTCCAC	540
AATCTGTTTT AGCAAGCGGC TTACAAGGAA TGGATGTAGT ACACGGCACA GCCACTATGC	600
AAGTAGATGG TAATAAAACC ATTATCCGCA ACAGTGTGTA CGATATCATT AATTGGAAAC	660
AATTTAACAT CGACCAAAAT GAAATGGTGC AGTTTTTACA AGAAAACAAC AACTCGCGCG	720
TATTCAACCG TGTTACATCT AACCAAATCT CCCAATTAAA AGGGATTTTA GATTCTAACG	780
GACAAGTCTT TTTAATCAAC CCAATGGTA TCACAATAGG TAAAGACGCA ATTATTAACA	840
CTAATGGCTT TACGGCTTCT ACGCTAGACA TTTCTAACGA AAACATCAAG GCGGTAATT	900
TCACCTTCGA GCAACCAAAA GATAAGCGC TCGCTGAAAT TGTGAATCAC GGTTTAATTA	960
CTGTCGGTAA AGACGGCAGT GTAATCTTA TTGGTGGCAA AGTGAAAAAC GAGGGTGTGA	1020
TTAGCGTAAA TGGTGGCAGC ATTTCTTTAC TCGCAGGGCA AAAAATCACC ATCAGCGATA	1080
TAATAAACCC AACCATTAAT TACAGCATTG CCGCGCTGA AAATGAAGCG GTCAATCTGG	1140
GCGATATTTT TGCCAAAGCG GGTAAACATTA ATGTCCGTGC TGCCACTATT CGAAACCAAG	1200
GTAACCTTTC TGCTGATTCT GTAAGCAAAG ATAAAGCGG CAATATTGTT CTTTCCGCCA	1260
AAGAGGGTGA AGCGGAAATT GCGGTGTAA TTTCCGCTCA AAATCAGCAA GCTAAGGGCG	1320
GCAAGCTGAT GATTACAGCG GATAAAGTCA CATTAAAAAC AGGTGCAAGT ATCGACCTTT	1380
CAGGTAAAGA AGGGGGAGAA ACTTACCTTG GCGGTGACGA GCGCGCGGAA GGTAAAAAGG	1440
GCATTCAATT AGCAAGAGAA ACCTCTTTAG AAAAAGGCTC AACCATCAAT GTATCAGGCA	1500
AAGAAAAAGG CGGACGCGCT ATTGTGTGGG GCGATATTGC GTTAATTGAC GGCAATATTA	1560
ACGCTCAAGG TAGTGGTGAT ATCGCTAAAA CCGGTGGTTT TGTGGAGACG TCGGGGCATG	1620
ATTTATTCAT CAAAGACAAAT GCAATTGTTG ACGCCAAGA GTGGTTGTTA GACCCGGATA	1680
ATGTATCTAT TAATGCAGAA ACAGCAGGAC GCAGCAATAC TTCAGAAGAC GATGAATACA	1740
CGGGATCCGG GAATAGTGCC AGCACCCCAA AACGAAACAA AGAAAAGACA ACATTAAACAA	1800
ACACAACCTT TGAGAGTATA CTAaaaaaag GTACCTTTGT TAACATCACT GCTAATCAAC	1860
GCACTATATG CAATAGCTCC ATTAATTTAT CCAATGSCAG CTTAACTCTT TGGAGTGAGG	1920
GTGCGAGCGG TGGCGGCGTT GAGATTAAAC ACGATATTAC CACCGGTGAT GATACCAAGG	1980
GTGCAAACTT AACAAATTAC TCAGGCGGCT GGGTTGATGT TCATAAAAAT ATCTCACTCG	2040
GGGCGCAAGG TAACATAAAC ATTACAGCTA AACAGATAT CGCCTTTGAG AAAGGAAGCA	2100
ACCAAGTCAT TACAGGTCAA GGGACTATTA CCTCAGGCAA TCAAAAAGGT TTTGATTTA	2160
ATAATGTCTC TCTAACCGCG ACTGGCAGCG GACTGCAATT CACCATAAA AGAACCAATA	2220

AATACGCTAT CACAAATAAA TTTGAAGGGA CTTTAAATAT TTCAGGGA AAA GTGAACATCT	2280
CAATGGTTTT ACCTAAAAAT GAAAGTGGAT ATGATAAAT CAAAGGACGC ACTTACTGGA	2340
ATTTAACCTC CTTAAATGTT TCCGAGAGTG GCGAGTTTAA CCTCACTATT GACTCCAGAG	2400
GAAGCGATAG TGCAGGCACA CTTACCCAGC CTTATAATTT AAACGGTATA TCATTCAACA	2460
AAGACACTAC CTTTAAATGTT GAACGAAATG CAAGAGTCAA CTTTGACATC AAGGCACCAA	2520
TAGGGATAAA TAAGTATTCT AGTTTGAATT ACGCATCATT TAATGGAAAC ATTTCAAGTTT	2580
CGGGAGGGGG GAGTGTGTAT TTCACACTTC TCGCTCATC CTCTAACGTC CAAACCCCG	2640
GTGTAGTTAT AAATCTTAAA TACTTTAATG TTTCACAGS GTCAAGTTTA AGATTTAAAA	2700
CTTCAGGCTC AACAAAAACT GGCTTCTCAA TAGAGAAAGA TTTAACTTTA AATGCCACCG	2760
GAGGCAACAT AACACTTTTG CAAGTTGAAG GCACCGATGG AATGATTGGT AAAGGCATTG	2820
TAGCCAAAAA AACATAACC TTTGAAGGAG GTAACATCAC CTTTGGCTCC AGGAAAGCCG	2880
TAACGAAAT CGAAGGCAAT GTTACTATCA ATAACAACGC TAACGTCCT CTTATCGGTT	2940
CGGATTTTGA CAACCATCAA AAACCTTTAA CTATTAAAA AGATGTCATC ATTAATAGCG	3000
GCAACCTTAC CGCTGGAGGC AATATTGTCA ATATAGCCGG AAATCTTACC GTTGAAAGTA	3060
ACGCTAATTT CAAAGCTATC ACAAATTTCA CTTTAAATGT AGGCGGCTTG TTTGACAA	3120
AAGGCAATTC AATATTTTCC ATTGCCAAG GAGGGGCTCG CTTTAAAGAC ATTGATAAT	3180
CCAAGAATTT AAGCATCACC ACCAACTCCA GCTCCACTTA CCGCACTATT ATAAGCGGCA	3240
ATATAACCAA TAAAAACGGT GATTTAAATA TTACGAACGA AGGTAGTGAT ACTGAAATGC	3300
AAATTGGCGG CGATGTCTCG CAAAAAGAAG GTAATCTCAC GATTTCTTCT GACAAAAATCA	3360
ATATTACCAA ACAGATAACA ATCAAGGCAG GTGTTGATGG GGAGAATTCC GATTGAGACG	3420
CGACAAACAA TGCCAATCTA ACCATTAAAA CCAAGAATT GAAATTAAAG CAAGACCTAA	3480
ATATTTCAGG TTTCAATAAA GCAGAGATTA CAGCTAAAGA TGGTAGTGAT TTAACATTG	3540
GTAACACCAA TAGTGCTGAT GGTACTAATG CAAAAAAGT AACCTTTAAC CAGGTTAAAG	3600
ATTCAAAAT CTCTGCTGAC GGTCACAAGG TGCACTACA CAGCAAAGTG GAAACATCCG	3660
GTAGTAATAA CAACACTGAA GATAGCAGTG ACAAATAATGC CGGCTTAATC ATCGATGCAA	3720
AAAATGTAAC AGTAACAAC AATATTACTT CTCACAAAGC AGTGAGCATC TCTGCGACAA	3780
GTGAGAAAT TACCATAAA ACAGGTACAA CCATTAAAGC AACCCTGGT AACGTGGAGA	3840
TAACCGCTCA AACAGGTAGT ATCCTAGGTG GAATTGAGTC CAGCTCTGGC TCTGTAACAC	3900
TTACTGCAAC CGAGGGCGCT CTGCTGTAA GCAATATTTC GGGCAACACC GTTACTGTTA	3960
CTGCAAAATG CGGTGCATTA ACCACTTTGG CAGGCTCTAC AATTAAAGGA ACCGAGAGTG	4020
TAACCACTTC AAGTCAATCA GCGATATCG GCGGTACGAT TTCTGGTGGC ACAGTAGAGG	4080
TTAAAGCAAC CGAAAGTTTA ACCACTCAAT CCAATTCAAA AATTAAAGCA ACAACAGGCG	4140
AGGCTAACGT AACAAAGTGA ACAGGTACAA TTGGTGGTAC GATTTCCGGT AATACGGTAA	4200
ATGTTCACGC AACCGCTGCG GATTTAACAG TTGGGAATGG CGCAGAAATT AATGCCACAG	4260

AAGGAGCTGC AACCTTAACT ACATCATCGG GCAAATTAAC TACCGAAGCT AGTTCACACA 4320
 TTACTTTCAGC CAAGGGTCAG GTAAATCTTT CAGCTCAGGA TGGTAGCGTT GCAGGAAGTA 4380
 TTAATGCCCG CAATGTGACA CTAAATACTA CAGGCACTTT AACTACCGTG AAGGGTTCAA 4440
 ACATTATATGC AACCAGCGGT ACCTTGGTTA TTAACGCAAA AGACGCTGAG CTAATGGCGG 4500
 CAGCATTTGGG TAACCACACA GTGGTAAATG CAACCAACGC AAATGGCTCC GGCACGCTAA 4560
 TCGCGACAAC CTCAAGCAGA GTGAACATCA CTGGGGATTT AATCACATAA AATGGATTAA 4620
 ATATCATTTTC AAAAAACGGT ATAAACACCG TACTGTTTAA AGCGCTTAAA ATTGATGTGA 4680
 AATCATTTCA ACCGGGTATA GCAAGCGTAG ATGAAGTAAT TGAAGCGAAA CGCATCCTTG 4740
 AGAAGGTAAA AGATTATCT GATGAAGAAA GAGAAGCGTT AGCTAAACTT GGAGTAAGTG 4800
 CTGTACGTTT TATTGAGCCA AATAATACAA TTACAGTCGA TACACAAAAT GAATTGCAA 4860
 CCAGACCATT AAGTCGAATA GTGATTCTG AAGGCAGGGC GTGTTTCTCA AACAGTAGT 4920
 GCGCGACGGT GTGCGTTAAT ATCGTGTATA ACGGGCGGTA GCGGTCAGTA ATTGACAAGG 4980
 TAGATTTTAT CCTGCAATGA AGTCATTTTA TTTTCGTATT ATTTACTGTG TGGGTTAAAG 5040
 TTCAGTACGG GCTTTACCCA TCTTGTAATA AATTACGGAG AATACAATAA AGTATTTTAA 5100
 ACAGGTTATT ATTATG 5116

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1536 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu
 1 5 10 15
 Val Ala Val Ser Glu Leu Ala Arg Gly Cys Asp His Ser Thr Glu Lys
 20 25 30
 Gly Ser Glu Lys Pro Ala Arg Met Lys Val Arg His Leu Ala Leu Lys
 35 40 45
 Pro Leu Ser Ala Met Leu Leu Ser Leu Gly Val Thr Ser Ile Pro Gln
 50 55 60
 Ser Val Leu Ala Ser Gly Leu Gln Gly Met Asp Val Val His Gly Thr
 65 70 75 80
 Ala Thr Met Gln Val Asp Gly Asn Lys Thr Ile Ile Arg Asn Ser Val
 85 90 95
 Asp Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met
 100 105 110
 Val Gln Phe Leu Gln Glu Asn Asn Asn Ser Ala Val Phe Asn Arg Val
 115 120 125

Thr Ser Asn Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly
 130 135 140
 Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
 145 150 155 160
 Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
 165 170 175
 Glu Asn Ile Lys Ala Arg Asn Phe Thr Phe Glu Gln Thr Lys Asp Lys
 180 185 190
 Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
 195 200 205
 Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
 210 215 220
 Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
 225 230 235 240
 Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
 245 250 255
 Glu Asn Glu Ala Val Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn
 260 265 270
 Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Gln Gly Lys Leu Ser Ala
 275 280 285
 Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
 290 295 300
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
 305 310 315 320
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
 325 330 335
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
 340 345 350
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala
 355 360 365
 Lys Lys Thr Ser Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
 370 375 380
 Glu Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
 385 390 395 400
 Gly Asn Ile Asn Ala Gln Gly Ser Gly Asp Ile Ala Lys Thr Gly Gly
 405 410 415
 Phe Val Glu Thr Ser Gly His Asp Leu Phe Ile Lys Asp Asn Ala Ile
 420 425 430
 Val Asp Ala Lys Glu Trp Leu Leu Asp Phe Asp Asn Val Ser Ile Asn
 435 440 445
 Ala Glu Thr Ala Gly Arg Ser Asn Thr Ser Glu Asp Asp Glu Tyr Thr
 450 455 460
 Gly Ser Gly Asn Ser Ala Ser Thr Pro Lys Arg Asn Lys Glu Lys Thr
 465 470 475 480

Thr Leu Thr Asn Thr Thr Leu Glu Ser Ile Leu Lys Lys Gly Thr Phe
 485 490 495
 Val Asn Ile Thr Ala Asn Gln Arg Ile Tyr Val Asn Ser Ser Ile Asn
 500 505 510
 Leu Ser Asn Gly Ser Leu Thr Leu Trp Ser Glu Gly Arg Ser Gly Gly
 515 520 525
 Gly Val Glu Ile Asn Asn Asp Ile Thr Thr Gly Asp Asp Thr Arg Gly
 530 535 540
 Ala Asn Leu Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn
 545 550 555 560
 Ile Ser Leu Gly Ala Gln Gly Asn Ile Asn Ile Thr Ala Lys Gln Asp
 565 570 575
 Ile Ala Phe Glu Lys Gly Ser Asn Gln Val Ile Thr Gly Gln Gly Thr
 580 585 590
 Ile Thr Ser Gly Asn Gln Lys Gly Phe Arg Phe Asn Asn Val Ser Leu
 595 600 605
 Asn Gly Thr Gly Ser Gly Leu Gln Phe Thr Thr Lys Arg Thr Asn Lys
 610 615 620
 Tyr Ala Ile Thr Asn Lys Phe Glu Gly Thr Leu Asn Ile Ser Gly Lys
 625 630 635 640
 Val Asn Ile Ser Met Val Leu Pro Lys Asn Glu Ser Gly Tyr Asp Lys
 645 650 655
 Phe Lys Gly Arg Thr Tyr Trp Asn Leu Thr Ser Leu Asn Val Ser Glu
 660 665 670
 Ser Gly Glu Phe Asn Leu Thr Ile Asp Ser Arg Gly Ser Asp Ser Ala
 675 680 685
 Gly Thr Leu Thr Gln Pro Tyr Asn Leu Asn Gly Ile Ser Phe Asn Lys
 690 695 700
 Asp Thr Thr Phe Asn Val Glu Arg Asn Ala Arg Val Asn Phe Asp Ile
 705 710 715 720
 Lys Ala Pro Ile Gly Ile Asn Lys Tyr Ser Ser Leu Asn Tyr Ala Ser
 725 730 735
 Phe Asn Gly Asn Ile Ser Val Ser Gly Gly Ser Val Asp Phe Thr
 740 745 750
 Leu Leu Ala Ser Ser Ser Asn Val Gln Thr Pro Gly Val Val Ile Asn
 755 760 765
 Ser Lys Tyr Phe Asn Val Ser Thr Gly Ser Ser Leu Arg Phe Lys Thr
 770 775 780
 Ser Gly Ser Thr Lys Thr Gly Phe Ser Ile Glu Lys Asp Leu Thr Leu
 785 790 795 800
 Asn Ala Thr Gly Gly Asn Ile Thr Leu Leu Gln Val Glu Gly Thr Asp
 805 810 815
 Gly Met Ile Gly Lys Gly Ile Val Ala Lys Lys Asn Ile Thr Phe Glu
 820 825 830

Gly Gly Asn Ile Thr Phe Gly Ser Arg Lys Ala Val Thr Glu Ile Glu
 835 840 845
 Gly Asn Val Thr Ile Asn Asn Asn Ala Asn Val Thr Leu Ile Gly Ser
 850 855 860
 Asp Phe Asp Asn His Gln Lys Pro Leu Thr Ile Lys Lys Asp Val Ile
 865 870 875 880
 Ile Asn Ser Gly Asn Leu Thr Ala Gly Gly Asn Ile Val Asn Ile Ala
 885 890 895
 Gly Asn Leu Thr Val Glu Ser Asn Ala Asn Phe Lys Ala Ile Thr Asn
 900 905 910
 Phe Thr Phe Asn Val Gly Gly Leu Phe Asp Asn Lys Gly Asn Ser Asn
 915 920 925
 Ile Ser Ile Ala Lys Gly Gly Ala Arg Phe Lys Asp Ile Asp Asn Ser
 930 935 940
 Lys Asn Leu Ser Ile Thr Thr Asn Ser Ser Ser Thr Tyr Arg Thr Ile
 945 950 955 960
 Ile Ser Gly Asn Ile Thr Asn Lys Asn Gly Asp Leu Asn Ile Thr Asn
 965 970 975
 Glu Gly Ser Asp Thr Glu Met Gln Ile Gly Gly Asp Val Ser Gln Lys
 980 985 990
 Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Ile Asn Ile Thr Lys Gln
 995 1000 1005
 Ile Thr Ile Lys Ala Gly Val Asp Gly Glu Asn Ser Asp Ser Asp Ala
 1010 1015 1020
 Thr Asn Asn Ala Asn Leu Thr Ile Lys Thr Lys Glu Leu Lys Leu Thr
 1025 1030 1035 1040
 Gln Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr Ala Lys
 1045 1050 1055
 Asp Gly Ser Asp Leu Thr Ile Gly Asn Thr Asn Ser Ala Asp Gly Thr
 1060 1065 1070
 Asn Ala Lys Lys Val Thr Phe Asn Gln Val Lys Asp Ser Lys Ile Ser
 1075 1080 1085
 Ala Asp Gly His Lys Val Thr Leu His Ser Lys Val Glu Thr Ser Gly
 1090 1095 1100
 Ser Asn Asn Asn Thr Glu Asp Ser Ser Asp Asn Asn Ala Gly Leu Thr
 1105 1110 1115 1120
 Ile Asp Ala Lys Asn Val Thr Val Asn Asn Asn Ile Thr Ser His Lys
 1125 1130 1135
 Ala Val Ser Ile Ser Ala Thr Ser Gly Glu Ile Thr Thr Lys Thr Gly
 1140 1145 1150
 Thr Thr Ile Asn Ala Thr Thr Gly Asn Val Glu Ile Thr Ala Gln Thr
 1155 1160 1165
 Gly Ser Ile Leu Gly Gly Ile Glu Ser Ser Ser Gly Ser Val Thr Leu
 1170 1175 1180

Thr Ala Thr Glu Gly Ala Leu Ala Val Ser Asn Ile Ser Gly Asn Thr
 1185 1190 1195 1200
 Val Thr Val Thr Ala Asn Ser Gly Ala Leu Thr Thr Leu Ala Gly Ser
 1205 1210 1215
 Thr Ile Lys Gly Thr Glu Ser Val Thr Thr Ser Ser Gln Ser Gly Asp
 1220 1225 1230
 Ile Gly Gly Thr Ile Ser Gly Gly Thr Val Glu Val Lys Ala Thr Glu
 1235 1240 1245
 Ser Leu Thr Thr Gln Ser Asn Ser Lys Ile Lys Ala Thr Thr Gly Glu
 1250 1255 1260
 Ala Asn Val Thr Ser Ala Thr Gly Thr Ile Gly Gly Thr Ile Ser Gly
 1265 1270 1275 1280
 Asn Thr Val Asn Val Thr Ala Asn Ala Gly Asp Leu Thr Val Gly Asn
 1285 1290 1295
 Gly Ala Glu Ile Asn Ala Thr Glu Gly Ala Ala Thr Leu Thr Thr Ser
 1300 1305 1310
 Ser Gly Lys Leu Thr Thr Glu Ala Ser Ser His Ile Thr Ser Ala Lys
 1315 1320 1325
 Gly Gln Val Asn Leu Ser Ala Gln Asp Gly Ser Val Ala Gly Ser Ile
 1330 1335 1340
 Asn Ala Ala Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Val
 1345 1350 1355 1360
 Lys Gly Ser Asn Ile Asn Ala Thr Ser Gly Thr Leu Val Ile Asn Ala
 1365 1370 1375
 Lys Asp Ala Glu Leu Asn Gly Ala Ala Leu Gly Asn His Thr Val Val
 1380 1385 1390
 Asn Ala Thr Asn Ala Asn Gly Ser Gly Ser Val Ile Ala Thr Thr Ser
 1395 1400 1405
 Ser Arg Val Asn Ile Thr Gly Asp Leu Ile Thr Ile Asn Gly Leu Asn
 1410 1415 1420
 Ile Ile Ser Lys Asn Gly Ile Asn Thr Val Leu Leu Lys Gly Val Lys
 1425 1430 1435 1440
 Ile Asp Val Lys Tyr Ile Gln Pro Gly Ile Ala Ser Val Asp Glu Val
 1445 1450 1455
 Ile Glu Ala Lys Arg Ile Leu Glu Lys Val Lys Asp Leu Ser Asp Glu
 1460 1465 1470
 Glu Arg Glu Ala Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Ile
 1475 1480 1485
 Glu Pro Asn Asn Thr Ile Thr Val Asp Thr Gln Asn Glu Phe Ala Thr
 1490 1495 1500
 Arg Pro Leu Ser Arg Ile Val Ile Ser Glu Gly Arg Ala Cys Phe Ser
 1505 1510 1515 1520
 Asn Ser Asp Gly Ala Thr Val Cys Val Asn Ile Ala Asp Asn Gly Arg
 1525 1530 1535

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4937 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TAAATATACA AGATAATATAA AATAAATCAA GATTTTGTG ATGACAAACA ACAATTACAA	60
CACCTTTTT GCAGTCTATA TGCAAAATATT TTAATAAAT AGTATAAATC CGCCATATAA	120
AATGGTATAA TCTTTCATCT TTCATCTTTA ATCTTTCATC TTTCATCTTT CATCTTTCAT	180
CTTTCATCTT TCATCTTTCA TCTTTCATCT TTCATCTTTC ATCTTTCATC TTTTCATCTT	240
CACATGAAAT GATGAACCGA GGAAGGGGAG GGAGGGGCAA GAATGAAGAG GGAGCTGAAC	300
GAACGCAAAAT GATAAAGTAA TTTAATTGTT CACTAACCT TAGGAGAAAA TATGAACAAAG	360
ATATATCGTC TCAAAATTCAG CAAACGCGTG AATGCTTTGG TTGCTGTGTC TGAATTGGCA	420
CGGGGTGTG ACCATTCCAC AGAAAAAGGC TTCCGCTATG TTACTATCTT TAGTGTAAC	480
CACCTAGCGT TAAAGCCACT TTCCGCTATG TTACTATCTT TAGTGTAAC ATCTATTCCA	540
CAATCTGTTT TAGCAAGCGG CTTACAAGGA ATGGATGTAG TACACGGCAC AGCCACTATG	600
CAAGTAGATG GTAATAAAAC CATTATCCGC AACAGTGTG ACGCTATCAT TAATTGAAAA	660
CAATTTAACA TCGACCAAAA TGAAATGGTG CAGTTTTTAC AAGAAAACAA CAACTCCGCC	720
GTATTC AACG GTGTACATC TAACCAAATC TCCCAATTA AAGGGATTTT AGATTCTAAC	780
GGACAAGTCT TTTTATACAA CCCAAATGGT ATCACAATAG GTAAAGACGC AATTATTAAAC	840
ACTAATGGCT TTACGCGCTC TACGCTAGAC ATTTCTAACG AAAACATCAA GCGCGTAAT	900
TTACCTTCG AGCAAAACCA AGATAAAGCG CTCGCTGAAA TTGTGAATCA CGGTTTAATT	960
ACTGTCGGTA AAGACGGCAG TGTAATCTT ATTGTGGCA AAGTAAAAA CGAGGGGTGTG	1020
ATTAGCGTAA ATGTGGCAG CATTTCTTTA CTCGCAGGCG AAAAATACAT CATCAGCGAT	1080
ATAATAAACC CAACATTAC TTACAGCATT GCCGCGCGTG AAAATGAAGC GGTCAATCTG	1140
GGCGATATTT TTGCCAAAGG CGGTAAACATT AATGTCGGTG CTGCCACTAT TCGAAACCAA	1200
GGTAAACTTT CTGCTGATTC TGTAAGCAA GATAAAGCG GCAATATTGT TCTTTCGCCC	1260
AAAGAGGGTG AAGCGGAAAT TGGCGGTGTA ATTTCCGCTC AAAATCAGCA AGCTAAAGGC	1320
GGCAAGCTGA TGATTACAGG CGATAAAGTC ACATTAAAAA CAGGTGCAGT TATCGACCTT	1380
TCAGGTAAG AAGGGGAGAG AACTTACCTT GCGCGTGACG AGCGCGGCGA AGGTAAAAAC	1440
GGCATTCAAT TAGCAAAGAA AACCTCTTTA GAAAAGGCT CAACCATCAA TGTATCAGGC	1500
AAAGAAAAAG GCGACGCGC TATTGTGTGG GCGATATTG CGTTAATTGA CGGCAATATT	1560
AACGCTCAAG GTAGTGTGTA TATCGCTAAA ACGGTGTGTT TTGTGGAGAC ATCGGGGCAT	1620
TATTTATCCA TTGACAGCAA TGCAATTGTT AAAACAAAAG AGTGGTTGCT AGACCTGAT	1680

GATGTAACAA	TTGAAGCCGA	AGACCCCCTT	CGCAATAATA	CGGTATAAA	TGATGAATTG	1740
CCACACAGCA	CGGTGAAGC	AAGCGACCCT	AAAAAAAATA	CGRAACTCAA	AACAACGCTA	1800
ACCAATACAA	CTATTTCAAA	TTATCTGAAA	AACGCCTGGA	CAATGAATAT	AACGGCATCA	1860
AGAAAACTTA	CGTTAATAG	CTCAATCAAC	ATCGGAAGCA	ACTCCCACIT	TAATTCTCCAT	1920
AGTAAAGGTC	AGCGTGGCGG	AGCGGTTGAG	ATTGATGGAG	ATATTACTTC	TAAAGGCGGA	1980
AATTTAACCA	TTTATTCTGG	CGGATGGGTT	GATGTTTCATA	AAAATATTAC	GCTTGATCAG	2040
GGTTTTTTAA	ATATTACCGC	CGCTTCCGTA	GCTTTTGAAG	GTGGAATAAA	CAAAGCACGC	2100
GACGCGGCAA	ATGCTAAAA	TGTCGCCAG	GGCACTGTAA	CCATTACAGG	AGAGGGAAAA	2160
GATTTACAGG	CTAACACGCT	ATCTTTAAAC	GGAACGGGTA	AAGGTCTGAA	TATCATTTC	2220
TCAGTGAATA	ATTTAACCCA	CAATCTTAGT	GGCACAATTA	ACATATCTGG	GAATATAACA	2280
ATTAACCAAA	CTACGAGAAA	GAACACCTCG	TATTGGCAAA	CCAGCCATGA	TTCGCACTGG	2340
AACGTCAGTG	CTCTTAATCT	AGAGACAGGC	GCAAAATTTA	CCTTTATTAA	ATACATTTC	2400
AGCAATAGCA	AAGGCTTAAC	AACACAGTAT	AGAAGCTCTG	CAGGGGTGAA	TTTAAACGGC	2460
GTAATGGCA	ACATGTCATT	CAATCTCAAA	GAAGGAGCGA	AAGTTAATTT	CRAATTAATA	2520
CCAAACGAGA	ACATGAACAC	AAGCAAACCT	TTACCAATTC	GGTTTTTAGC	CAATATCACA	2580
GCCACTGGTG	GGGGCTCTGT	TTTTTTTGAT	ATATATGCCA	ACCATTCTGG	CAGAGGGGCT	2640
GAGTTAAAA	TGAGTGAAT	TAATATCTCT	AACGGCGCTA	ATTTTACCTT	AAATCCCAT	2700
GTTCGCGGCG	ATGACGCTTT	TAAAAACAAC	AAAGACTTAA	CCATAAATGC	AACCAATTCA	2760
AATTTACAGC	TCAGACAGAC	GAAAGATGAT	TTTTATGACG	GGTACGCACG	CAATGCCATC	2820
AATTCACCT	ACAACATATC	CATTCTGGGC	GGTAATGTCA	CCCTTGGTGG	ACAAAACCTCA	2880
AGCAGCAGCA	TTACGGGGAA	TATTACTATC	GAGAAAGCAG	CAATGTTCAC	GCTAGAAGCC	2940
AATAACGCCC	CTAATCAGCA	AAACATAAGG	GATAGAGTTA	TAAACTTGG	CAGCTTGCTC	3000
GTTAATGGGA	GTTTAAGTTT	AACTGGCGAA	AATGCAGATA	TTAAAGGCAA	TCTCACTATT	3060
TCAGAAAGCG	CCACTTTTAA	AGGAAAGACT	AGAGATACCC	TAAATATCAC	CGGCAATTTT	3120
ACCAATAATG	GCACCTGCGA	AATTAATATA	ACACAAGGAG	TGGTAAACT	TGGCAATGTT	3180
ACCAATGATG	GTGATTTAAA	CATTACCACT	CACGCTAAAC	GCAACCAAAG	AAGCATCATC	3240
GGCGGAGATA	TAATCAACAA	AAAAGGAAGC	TTAAATATTA	CAGACAGTAA	TAATGATGCT	3300
GAAATCCAAA	TTGGCGGCAA	TATCTCGCAA	AAAGAAGGCA	ACCTCACGAT	TTCTTCCGAT	3360
AAATTAATA	TCACCAAACA	GATAACAATC	AAAAAGGGTA	TTGATGGAGA	GGACTCTAGT	3420
TCAGATGCGA	CAAGTAATGC	CAACCTAACT	ATTAAAACCA	AAGAATTGAA	ATTGACAGAA	3480
GACCTAAGTA	TTTCAGGTTT	CAATAAAGCA	GAGATTACAG	CCAAGATGAG	TAGAGATTTA	3540
ACTATTGGCA	ACAGTAATGA	CGGTAACAGC	GGTGCCGAAG	CCAAACACAGT	AACTTTTAAAC	3600
AATGTTAAAG	ATTCAAAAA	CTCTGCTGAC	GGTCACATG	TGACACTAAA	TAGCAAAGTG	3660
AAAACATCTA	GCAGCAATGG	CGGACGTGAA	AGCAATAGCG	ACAACGATAC	CGGCTTAAC	3720

ATTACTGCAA AAAATGTAGA AGTAAACAAA GATATTACTT CTCTCAAAAC AGTAAATATC 3780
 ACCGCGTCGG AAAAGGTTAC CACCACAGCA GGCTCGACCA TTAACGCAAC AAATGGCAAA 3840
 GCAAGTATTA CAACCAAAAC AGGTGATATC AGCGGTACGA TTTCGGGTAA CACGGTAAGT 3900
 GTTAGCGCGA CTGGTGATT AACCACTAAA TCCGGCTCAA AAATTGAAGC GAAATCGGGT 3960
 GAGGCTAATG TAACAAGTGC AACAGGTACA ATTGGCGGTA CAATTTCCGG TAATACGGTA 4020
 AATGTTACGG CAAACGCTGG CGATTTAACA GTTGGGAATG GCGCAGAAAT TAATGCGACA 4080
 GAAGGAGCTG CAACCTTAAC CGCAACAGGG AATACCTTGA CTAAGTGAAGC CGGTCTTAGC 4140
 ATCACTTCAA CTAAGGGTCA GGTAGACCTC TTGGCTCAGA ATGGTAGCAT CGCAGGAAGC 4200
 ATTAATGCTG CTAATGTGAC ATTAATACT ACAGGCACCT TAACCACCGT GGCAGGCTCG 4260
 GATATTAAG CAACCAGCGG CACCTTGGTT ATTAACGCAA AAGATGCTAA GCTAAATGGT 4320
 GATGCATCAG GTGATAGTAC AGAAGTGAAT GCAGTCAACG CAAGCGGCTC TGGTAGTGTG 4380
 ACTGCGGCAA CCTCAAGCAG TGTGAATATC ACTGGGGATT TAAACACAGT AAATGGGTTA 4440
 AATATCATTT CGAAAGATGG TAGAAACACT GTGCGCTTAA GAGGCAAGGA AATTGAGGTG 4500
 AAATATATCC AGCCAGGTGT AGCAAGTGTGA GAAGAAGTAA TTGAAGCGAA ACGCGTCCTT 4560
 GAAAAAGTAA AAGATTATC TGATGAAGAA AGAGAAACAT TAGCTAAACT TGGTGTAAAT 4620
 GCTGTACGTT TTGTGAGGCC AAATAATACA ATTACAGTCA ATACACAAAA TGAATTTACA 4680
 ACCAGACCGT CAAGTCAAGT GATAATTTCT GAAGGTAAAG CGTGTCTCTC AAGTGGTAAT 4740
 GGCGCACGAG TATGTACCAA TGTGTGTGAC GATGGACAGC CGTAGTCAGT AATTGACAAG 4800
 GTAGATTTC TCTGCAATG AAGTCATTT ATTTTCGTAT TATTACTGT GTGGGTTAAA 4860
 GTTCAGTACG GGCTTTACCC ATCTTGTAAG AAATTACGGA GAATACAATA AAGTATTTTT 4920
 AACAGGTTAT TATTATG 4937

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1477 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu
 1 5 10 15
 Val Ala Val Ser Glu Leu Ala Arg Gly Cys Asp His Ser Thr Glu Lys
 20 25 30
 Gly Ser Glu Lys Pro Ala Arg Met Lys Val Arg His Leu Ala Leu Lys
 35 40 45
 Pro Leu Ser Ala Met Leu Leu Ser Leu Gly Val Thr Ser Ile Pro Gln
 50 55 60

Ser Val Leu Ala Ser Gly Leu Gln Gly Met Asp Val Val His Gly Thr
 65 70 75 80
 Ala Thr Met Gln Val Asp Gly Asn Lys Thr Ile Ile Arg Asn Ser Val
 85 90 95
 Asp Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met
 100 105 110
 Val Gln Phe Leu Gln Glu Asn Asn Asn Ser Ala Val Phe Asn Arg Val
 115 120 125
 Thr Ser Asn Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly
 130 135 140
 Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
 145 150 155 160
 Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
 165 170 175
 Glu Asn Ile Lys Ala Arg Asn Phe Thr Phe Glu Gln Thr Lys Asp Lys
 180 185 190
 Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
 195 200 205
 Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
 210 215 220
 Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
 225 230 235 240
 Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
 245 250 255
 Glu Asn Glu Ala Val Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn
 260 265 270
 Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Gln Gly Lys Leu Ser Ala
 275 280 285
 Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
 290 295 300
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
 305 310 315 320
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
 325 330 335
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
 340 345 350
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala
 355 360 365
 Lys Lys Thr Ser Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
 370 375 380
 Glu Lys Gly Gly Phe Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
 385 390 395 400
 Gly Asn Ile Asn Ala Gln Gly Ser Gly Asp Ile Ala Lys Thr Gly Gly
 405 410 415

Phe Val Glu Thr Ser Gly His Asp Leu Phe Ile Lys Asp Asn Ala Ile
 420 425 430
 Val Asp Ala Lys Glu Trp Leu Leu Asp Phe Asp Asn Val Ser Ile Asn
 435 440 445
 Ala Glu Asp Pro Leu Phe Asn Asn Thr Gly Ile Asn Asp Glu Phe Pro
 450 455 460
 Thr Gly Thr Gly Glu Ala Ser Asp Pro Lys Lys Asn Ser Glu Leu Lys
 465 470 475 480
 Thr Thr Leu Thr Asn Thr Thr Ile Ser Asn Tyr Leu Lys Asn Ala Trp
 485 490 495
 Thr Met Asn Ile Thr Ala Ser Arg Lys Leu Thr Val Asn Ser Ser Ile
 500 505 510
 Asn Ile Gly Ser Asn Ser His Leu Ile Leu His Ser Lys Gly Gln Arg
 515 520 525
 Gly Gly Glu Val Gln Ile Asp Gly Asp Ile Thr Ser Lys Gly Gly Asn
 530 535 540
 Leu Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn Ile Thr
 545 550 555 560
 Leu Asp Gln Gly Phe Leu Asn Ile Thr Ala Ala Ser Val Ala Phe Glu
 565 570 575
 Gly Gly Asn Asn Lys Ala Arg Asp Ala Ala Asn Ala Lys Ile Val Ala
 580 585 590
 Gln Gly Thr Val Thr Ile Thr Gly Glu Gly Lys Asp Phe Arg Ala Asn
 595 600 605
 Asn Val Ser Leu Asn Gly Thr Gly Lys Gly Leu Asn Ile Ile Ser Ser
 610 615 620
 Val Asn Asn Leu Thr His Asn Leu Ser Gly Thr Ile Asn Ile Ser Gly
 625 630 635 640
 Asn Ile Thr Ile Asn Gln Thr Thr Arg Lys Asn Thr Ser Tyr Trp Gln
 645 650 655
 Thr Ser His Asp Ser His Trp Asn Val Ser Ala Leu Asn Leu Glu Thr
 660 665 670
 Gly Ala Asn Phe Thr Phe Ile Lys Tyr Ile Ser Ser Asn Ser Lys Gly
 675 680 685
 Leu Thr Thr Gln Tyr Arg Ser Ser Ala Gly Val Asn Phe Asn Gly Val
 690 695 700
 Asn Gly Asn Met Ser Phe Asn Leu Lys Glu Gly Ala Lys Val Asn Phe
 705 710 715 720
 Lys Leu Lys Pro Asn Glu Asn Met Asn Thr Ser Lys Pro Leu Pro Ile
 725 730 735
 Arg Phe Leu Ala Asn Ile Thr Ala Thr Gly Gly Gly Ser Val Phe Phe
 740 745 750
 Asp Ile Tyr Ala Asn His Ser Gly Arg Gly Ala Glu Leu Lys Met Ser
 755 760 765

76.

Glu Ile Asn Ile Ser Asn Gly Ala Asn Phe Thr Leu Asn Ser His Val
 770 775 780
 Arg Gly Asp Asp Ala Phe Lys Ile Asn Lys Asp Leu Thr Ile Asn Ala
 785 790 795 800
 Thr Asn Ser Asn Phe Ser Leu Arg Gln Thr Lys Asp Asp Phe Tyr Asp
 805 810 815
 Gly Tyr Ala Arg Asn Ala Ile Asn Ser Thr Tyr Asn Ile Ser Ile Leu
 820 825 830
 Gly Gly Asn Val Thr Leu Gly Gly Gln Asn Ser Ser Ser Ile Thr
 835 840 845
 Gly Asn Ile Thr Ile Glu Lys Ala Ala Asn Val Thr Leu Glu Ala Asn
 850 855 860
 Asn Ala Pro Asn Gln Gln Asn Ile Arg Asp Arg Val Ile Lys Leu Gly
 865 870 875 880
 Ser Leu Leu Val Asn Gly Ser Leu Ser Leu Thr Gly Glu Asn Ala Asp
 885 890 895
 Ile Lys Gly Asn Leu Thr Ile Ser Glu Ser Ala Thr Phe Lys Gly Lys
 900 905 910
 Thr Arg Asp Thr Leu Asn Ile Thr Gly Asn Phe Thr Asn Asn Gly Thr
 915 920 925
 Ala Glu Ile Asn Ile Thr Gln Gly Val Val Lys Leu Gly Asn Val Thr
 930 935 940
 Asn Asp Gly Asp Leu Asn Ile Thr Thr His Ala Lys Arg Asn Gln Arg
 945 950 955 960
 Ser Ile Ile Gly Gly Asp Ile Ile Asn Lys Lys Gly Ser Leu Asn Ile
 965 970 975
 Thr Asp Ser Asn Asn Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser
 980 985 990
 Gln Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Ile Asn Ile Thr
 995 1000 1005
 Lys Gln Ile Thr Ile Lys Lys Gly Ile Asp Gly Glu Asp Ser Ser Ser
 1010 1015 1020
 Asp Ala Thr Ser Asn Ala Asn Leu Thr Ile Lys Thr Lys Glu Leu Lys
 1025 1030 1035 1040
 Leu Thr Glu Asp Leu Ser Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr
 1045 1050 1055
 Ala Lys Asp Gly Arg Asp Leu Thr Ile Gly Asn Ser Asn Asp Gly Asn
 1060 1065 1070
 Ser Gly Ala Glu Ala Lys Thr Val Thr Phe Asn Asn Val Lys Asp Ser
 1075 1080 1085
 Lys Ile Ser Ala Asp Gly His Asn Val Thr Leu Asn Ser Lys Val Lys
 1090 1095 1100
 Thr Ser Ser Ser Asn Gly Gly Arg Glu Ser Asn Ser Asp Asn Asp Thr
 1105 1110 1115 1120

Gly Leu Thr Ile Thr Ala Lys Asn Val Glu Val Asn Lys Asp Ile Thr
 1125 1130 1135
 Ser Leu Lys Thr Val Asn Ile Thr Ala Ser Glu Lys Val Thr Thr Thr
 1140 1145 1150
 Ala Gly Ser Thr Ile Asn Ala Thr Asn Gly Lys Ala Ser Ile Thr Thr
 1155 1160 1165
 Lys Thr Gly Asp Ile Ser Gly Thr Ile Ser Gly Asn Thr Val Ser Val
 1170 1175 1180
 Ser Ala Thr Val Asp Leu Thr Thr Lys Ser Gly Ser Lys Ile Glu Ala
 1185 1190 1195 1200
 Lys Ser Gly Glu Ala Asn Val Thr Ser Ala Thr Gly Thr Ile Gly Gly
 1205 1210 1215
 Thr Ile Ser Gly Asn Thr Val Asn Val Thr Ala Asn Ala Gly Asp Leu
 1220 1225 1230
 Thr Val Gly Asn Gly Ala Glu Ile Asn Ala Thr Glu Gly Ala Ala Thr
 1235 1240 1245
 Leu Thr Ala Thr Gly Asn Thr Leu Thr Thr Glu Ala Gly Ser Ser Ile
 1250 1255 1260
 Thr Ser Thr Lys Gly Gln Val Asp Leu Leu Ala Gln Asn Gly Ser Ile
 1265 1270 1275 1280
 Ala Gly Ser Ile Asn Ala Ala Asn Val Thr Leu Asn Thr Thr Gly Thr
 1285 1290 1295
 Leu Thr Thr Val Ala Gly Ser Asp Ile Lys Ala Thr Ser Gly Thr Leu
 1300 1305 1310
 Val Ile Asn Ala Lys Asp Ala Lys Leu Asn Gly Asp Ala Ser Gly Asp
 1315 1320 1325
 Ser Thr Glu Val Asn Ala Val Asn Ala Ser Gly Ser Gly Ser Val Thr
 1330 1335 1340
 Ala Ala Thr Ser Ser Ser Val Asn Ile Thr Gly Asp Leu Asn Thr Val
 1345 1350 1355 1360
 Asn Gly Leu Asn Ile Ile Ser Lys Asp Gly Arg Asn Thr Val Arg Leu
 1365 1370 1375
 Arg Gly Lys Glu Ile Glu Val Lys Tyr Ile Gln Pro Gly Val Ala Ser
 1380 1385 1390
 Val Glu Glu Val Ile Glu Ala Lys Arg Val Leu Glu Lys Val Lys Asp
 1395 1400 1405
 Leu Ser Asp Glu Glu Arg Glu Thr Leu Ala Lys Leu Gly Val Ser Ala
 1410 1415 1420
 Val Arg Phe Val Glu Pro Asn Asn Thr Ile Thr Val Asn Thr Gln Asn
 1425 1430 1435 1440
 Glu Phe Thr Thr Arg Pro Ser Ser Gln Val Ile Ile Ser Glu Gly Lys
 1445 1450 1455
 Ala Cys Phe Ser Ser Gly Asn Gly Ala Arg Val Cys Thr Asn Val Ala
 1460 1465 1470

Asp Asp Gly Gln Pro
1475

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9171 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ACAGCGTTCT CTTAATACTA GTACAAACCC ACAATAAAAT ATGACAAACA ACAATTACAA	60
CACCTTTTTT CGAGCTCTATA TGCAAAATATT TTAAAAATA GTATAAATCC GCCATATAAA	120
ATGGTATAAT CTTTCATCTT TCATCTTTCA TCCTTCATCT TTCATCTTTC ATCTTTCATC	180
TTTCATCTTT CATCTTTTCAT CTTTCATCTT TCATCTTTCA TCCTTCATCT TTCATCTTTC	240
ACATGAAATG ATGAACCGAG GGAAGGGGAGG GAGGGGCAAG AATGAAGAGG GAGCTGAACG	300
AACGCAAAAT ATAAAGTAAT TTAATTGTTC AACTAACCTT AGGAGAAAAT ATGAACAAGA	360
TATATCGTCT CAAATTCAGC AAACGCCTGA ATGCTTTGGT TGCTGTGTCT GAATTGGCAC	420
GGGGTGTGTA CCATTCCACA GAAAAAGGCA GCGAAAAACC TGCTCGCATG AAAGTCCGTC	480
ACTTAGCGTT AAAGCCACTT TCCGCTATGT TACTATCTTT AGGTGTAACA TCTATTCCAC	540
AATCTGTTTT AGCAAGCGGC TTACAAGGAA TGGATGTAGT ACACGGCACA GCCACTATGC	600
AAGTAGATGG TAATAAAACC ATTATCCGCA ACAGTGTTGA CGCTATCATT AATTGGAAAC	660
AATTTAACAT CGACCAAAAT GAAATGGTGC AGTTTTTACA AGAAAAACAAC AACTCCGCCG	720
TATTCAACCG TGTTACATCT AACCAAATCT CCCAATTAAA AGGGATTTTA GATTCTAACG	780
GACAAGTCTT TTATATCAAC CCAATGGTA TCACAATAGG TAAAGACGCA ATTATTAACA	840
CTAATGGCTT TACGGCTTCT ACGCTAGACA TTTCTAACGA AAACATCAAG GCGCGTAATT	900
TCACCTTCGA GCAAACCAA GATAAAGCGC TCGCTGAAAT TGTGAATCAC GGTTAATTA	960
CTGTGGTAA AGACGGCAGT GTAATCTTA TTGGTGGCAA AGTGAAAAAC GAGGGTGTGA	1020
TTAGCGTAAA TGGTGGCAGT ATTTCTTTAC TCGCAGGCA AAAAATCACC ATCAGCGATA	1080
TAATAAACCC AACCAATTACT TACAGCATTG CCGCGCCTGA AAATGAAGCG GTCAATCTGG	1140
GCGATATTTT TGCCAAAGGC GGTAACATTA ATGTCGTGTC TGCCACTATT CGAAACCAAG	1200
CTTTCCGCCA AAGAGGTGTA AGCGGAAATT GCGGGTGTAA TTTCCGCTCA AAATCAGCAA	1260
GCTAAGGCGC GCAAGCTGAT GATTACAGGC GATAAAGTCA CATTAAAAAC AGGTGCAGTT	1320
ATCGACCTTT CAGGTAAGA AGGGGGAGAA ACTTACCTTG GCGGTGACGA GCGCGGCGAA	1380
GGTAAAAACG GCATTCAATT AGCAAAGAAA ACCTCTTTAG AAAAAGGCTC AACCATCAAT	1440
GTATCAGGCA AAGAAAAAGG CGGACGCGCT ATTGTGTGGG GCGATATTGC GTTAATTGAC	1500

GGCAATATTA ACGCTCAAGG TAGTGGTGAT ATCGCTAAAA CCGGTGGTTT TGTGGAGACG	1560
TCGGGGCATG ATTTATTCAT CAAAGACAAT GCAATTGTTG ACGCCAAAGA GTGGTTGTTA	1620
GACCCGGATA ATGTATCTAT TAATGCAGAA ACAGCAGGAC GCAGCAATAC TTCAGAAGAC	1680
GATGAATACA CGGGATCCGG GAATAGTGCC AGCACCCCAA AACGAAACAA AGAAAAGACA	1740
ACATTAACAA ACACAACCTT TGAGAGTATA CTAAAAAAG GTACCTTTGT TAACATCACT	1800
GCTAATCAAC GCATCTATGT CAATAGCTCC ATTAATTAT CCAATGGCAG CTTAATCTTT	1860
TGGAGTGAGG GTCGGAGCGG TGGCGGCGTT GAGATTAACA ACGATATTAC CACCGGTGAT	1920
GATACCAGAG GTGCAAACTT AACAAATTAC TCAGCGCGCT GGGTTGATGT TCATAAAAAAT	1980
ATCTCACTCG GGGCGCAAGG TAACATAAAC ATTACAGCTA AACAAGATAT CGCCTTTGAG	2040
AAAGGAAGCA ACCAAGTCAT TACAGGTCAA GGGACTATTA CCTCAGGCAA TCAAAAAGGT	2100
TTTAGATTTA ATAATGTCTC TCTAAACGGC ACTGGCAGCG GACTGCAATT CACCACTAAA	2160
AGAACCAATA AATACGCTAT CACAAATAAA TTTGAAGGGA CTTTAAATAT TTCAGGGA	2220
GTGAACATCT CAATGGTTTT ACCTAAAAAT GAAAGTGGAT ATGATAAATT CAAAGGACGC	2280
ACTTACTGGA ATTTAACCTC GAAAGTGGAT ATGATAAATT CAAAGGACGC CCTCACTATT	2340
GACTCCAGAG GAAGCGATAG TGCAGGCACA CTTACCAGC CTTATAATT AAACGGTATA	2400
TCATTCAACA AAGCACTAC CTTTAAATGT GAACGAAATG CAAGAGTCAA CTTTGACATC	2460
AAGGCACCAA TAGGGATAAA TAAGTATTCT AGTTTGAATT ACGCATCATT TAATGGAAAC	2520
ATTTAGTTT CGGGAGGGGG GAGTGTGAT TTCACACTTC TCGCCTCATC CTCTAACGTC	2580
CAAACCCCG GTGTAGTTAT AAATTCTAAA TACTTTAATG TTTCAACAGG GTCAAGTTTA	2640
AGATTAAAA CTTCAGGCTC AACAAAACT GGCTTCTCAA TAGAGAAAGA TTTAACTTTA	2700
AATGCCACCG GAGGCAACAT AACACTTTTG CAAGTTGAAG GCACCGATGG AATGATTGGT	2760
AAAGGCATTG TAGCCAAAAA AACATATAAC TTTGAAGGAG GTAAGATGAG GTTTGGCTCC	2820
AGGAAAGCCG TAACAGAAAT CGAAGGCAAT GTTACTATCA ATAACAACGC TAACGTCACT	2880
CTTATCGGTT CGGATTTTGA CAACATCAA AACCTTTAA CTATTAAAA AGATGTCATC	2940
ATTAATAGCG GCAACCTTAC CGCTGGAGGC AATATTGTCA ATATAGCCGG AAATCTTACC	3000
GTGAAAGTA ACGCTAATTT CAAAGCTATC ACAAATTTCA CTTTAAATGT AGGCGGCTTG	3060
TTTGACACA AAGGCAATTC AAATATTTC ATTGCCAAG GAGGGGCTCG CTTTAAAGAC	3120
ATTGATAATT CCAAGAATTT AAGCATCACC ACCAATCCCA GCTCCACTTA CCGCACTATT	3180
ATAAGCGGCA ATATAACCA TAAAAACGGT GATTTAAATA TTACGAACGA AGGTAGTGAT	3240
ACTGAAATGC AAATTGGCGG CGATGTCTCG CAAAAAGAAG GTAATCTCAC GATTCTTCT	3300
GACAAATCA ATATTACCA ACAGATAACA ATCAAGGCAG GTGTTGATGG GGAGAATTCC	3360
GATTCAAGC CGACAAACAA TGCCAATCTA ACCATTAAAA CCAAGAATT GAAATTAAG	3420
CAAGACCTAA ATATTTCAGG TTTCAATAAA GCAGAGATTA CAGCTAAGA TGGTAGTGAT	3480
TTAATCTATG GTAACACCA TAGTGCTGAT GGTACTAATG CCAAAAAAGT AACCTTAAAC	3540

CAGGTTAAAG ATTCAAAAAAT CTCTGCTGAC GGTCAACAAG TGACACTACA CAGCAAAGTG	3600
GAAACATCCG GTAGTAATAA CAACACTGAA GATAGCAGTG ACAATAATGC CGGCTTAACT	3660
ATCGATGCAA AAAATGTAAAC AGTAAACAAC AATATTACTT CTCACAAAGC AGTGAGCATC	3720
TCTGCGACAA GTGGAGAAAA TACCACTAAA ACAGGTACAA CCATTAAACGC AACCACTGGT	3780
AACGTGGAGA TAACCGCTCA AACAGGTAGT ATCCTAGGTG GAATTGAGTC CAGCTCTGGC	3840
TCTGTAACAC TTACTGCAAC CGAGGGCGCT CTGTCTGTA GCAATATTTC GGGCAACACC	3900
GTTACTGTTA CTGCAATAG CGGTGCATTA ACCACTTTGG CAGGCTCTAC AATTAAAGGA	3960
ACCGAGAGTG TAACCACTTC AAGTCAATCA GGCATATCG CGGCTACGAT TTCTGGTGGC	4020
ACAGTAGAGG TTAAGCAAC CGAAAGTTTA ACCACTCAAT CCAATTCAAA AATTAAAGCA	4080
ACAACAGGCG AGGCTAACGT AACAAGTGCA ACAGGTACAA TTGGTGTGAC GATTTCGGGT	4140
AATACGGTAA ATGTTACGGC AAACGCTGGC GATTTAAACG TTGGGAATGG CGCAGAAATT	4200
AATGCGACAG AAGGAGCTGC AACCTTAACT ACATCATCGG GCAATTAAC TACCGAAGCT	4260
AGTTACACAC TTACTTCAGC CAAGGGTCAG GTAAATCTTT CAGCTCAGGA TGGTAGCGTT	4320
GCAGGAAGTA TTAATGCCGC CAATGTGACA CTAATACTA CAGGCACTTT AACTACCGTG	4380
AAGGGTTCAA ACATTAAATGC AACCAAGCGT ACCTTGGTTA TTAACGCAAA AGACGCTGAG	4440
CTAAATGGCG CAGCATTGGG TAACCACACA GTGGTAAATG CAACCAACGC AAATGGCTCC	4500
GGCAGCGTAA TCGCGACAAC CTCAGCAGA GTGAACATCA CTGGGGATTT AATCACAATA	4560
AATGGATTAA ATATCATTTC AAAAAACGGT ATAAACACCG TACTGTTAAA AGGCGTTAAA	4620
ATTGATGTGA AATACATTCA ACCGGGTATA GCAAGCGTAG ATGAAGTAAT TGAAGCGAAA	4680
CGCATCCTTG AGAAGGTAAA AGATTATCTT GATGAAGAAA GAGAAGCGTT AGCTAACTT	4740
GGCGTAAGTG CTGTACGTTT TATTGAGCCA AATAATACAA TTACAGTCGA TACACAAAAT	4800
GAATTTGCAA CCAGACCAAT AAGTCGAATA GTGATTCTGT AAGCAGGGC GTGTTTCTCA	4860
AACAGTGATG GCGCGACGGT TCGCGTTAAT ATCGCTGATA ACGGCGGGTA CGCGTCAGTA	4920
ATTGACAAGG TAGATTTCAT CCTGCAATGA AGTCATTTTA TTTTCGTATT ATTTACTGTG	4980
TGGGTTAAAG TTCAGTACGG GCTTTACCCA TCTTGTAATAA AATTACGGAG AATACAATAA	5040
AGTATTTTAA ACAGGTTATT ATTATGAAAA ATATAAAAA CAGATTAATAA CTCAGTGCAA	5100
TATCAGTATT GCTTGGCCTG GCTTCTTCAT CATGTATGTC AGAAGAACGG TTTTATAGTAA	5160
AAGGCTTTCA GTTATCTGGT GCACTTGAAA CTTTAAGTGA AGACGCCCAA CTGTCTGTAG	5220
CAAAATCTTT ATCTAAATAC CAAGGCTCGC AAACCTTAAAC AAACCTAAAA ACAGCACAGC	5280
TTGAATTACA GGCTGTGCTA GATAAGATTG AGCCAAATAA GTTTGATGTG ATATTGCCAC	5340
AACAAACCAT TACGATGGC AATATTATGT TTGAGCTAGT CTCGAAATCA GCGCGAGAAA	5400
GCCAAAGTTT TTATAAGGCG AGCCAGGGTT ATAGTGAAGA AAATATCGCT CGTAGCCTGC	5460
CATCTTTGAA ACAAGGAAAA GTGTATGAAG ATGCTCGTCA GTGGTTCGAT TTGCGTGAAT	5520
TCAATATGGC AAAAGAAAA CCACTTAAAG TCACTCGCGT GCATTACGAG TTAACCCTTA	5580

AAAACAAAC	CTCTGATTG	GTAGTTGCAG	GTTTTTCGCC	TTTTGGCAAA	ACGCGTAGCT	5640
TTGTTTCTA	TGATAATTTC	GGCGCAAGGG	AGTTTAACTA	TCAACGTGTA	AGTCTAGGTT	5700
TTGTTAATGC	CAATTTGACC	GGACATGATG	ATGTATTAAA	TCTAAACGCA	TTGACCAATG	5760
TAAAAGCACC	ATCAAAATCT	TATGCGGTAG	GCATAGGATA	TACTTATCCG	TTTTATGATA	5820
AACACCAATC	CTTAAGTCTT	TATACCAGCA	TGAGTTATGC	TGATTCTAAT	GATATCGAGC	5880
GCTTACCAAG	TGCGATTAAAT	CGTAAATTAT	CAAAAGGTCA	ATCTATCTCT	CGGAATCTGA	5940
AATGGAGTTA	TTATCTCCCG	ACATTTAACC	TTGGAATGGA	AGACCAGTTT	AAAATTAATT	6000
TAGGCTACAA	CTACCGCCAT	ATTATCAAAA	CATCCGAGTT	AAACACCTCG	GGTGCAACGA	6060
AGAAAAAATT	TGCAGTATCA	GGCGTAAGTG	CAGGCATTGA	TGGACATATC	CAATTTACCC	6120
CTAAAACAAT	CTTTAATATT	GATTTAAGTC	ATCAATTATTA	CGCGAGTAAA	TTACCAAGGCT	6180
CTTTTGGAAT	GGAGCGCATT	GGCGAAACAT	TTAATCGCAG	CTATCACATT	AGCACAGCCA	6240
GTTTAGGGTT	GAGTCAAGAG	TTTGCTCAAG	GTTGGCATT	TAGCAGTCAA	TTATCGGGTC	6300
AGTTTACTCT	ACAAGATATA	AGTAGCATAG	ATTTATTCTC	TGTAACAGGT	ACTTTAGGGG	6360
TCAGAGGCTT	TAAATACGGC	GGTGCAAGTG	GTGAGCGCGG	TCTTGTATGG	CGTAATGAAT	6420
TAAGTATGCC	AAATACACC	CGCTTTCAAA	TCAGCCCTTA	TGCGTTTTAT	GATGCAGGTC	6480
AGTCCGTTA	TAATAGCGAA	AATGCTAAAA	CTTACGGCGA	AGATATGCAC	ACGGTATCCT	6540
CTGCGGGTTT	AGGCATTAAA	ACCTCTCCTA	CACAAAACTT	AAGCTTAGAT	GCTTTTGTG	6600
CTCGTCGCT	TGCAATGCC	AATAGTGACA	ATTGGAATGG	CAACAAAAAA	CGCACAGACT	6660
CACCTACAAC	CTTCTGGGGT	AGATTAACT	TCAGTTTCTA	ACCCTGAAAT	TTAATCAACT	6720
GGTAAGCGTT	CCGCTACCA	GTTTATAACT	ATATGCTTTA	CCGCCAATT	TACAGTCTAT	6780
ACGCAACCTT	GTTTTTCATCC	TTATATATCA	AACAACTAA	GCAACCAAG	CAACCAAGC	6840
AAACCAAGCA	AACCAAGCAA	ACCAAGCRAA	CCAAGCAAAC	CAAGCAAACC	AAGCAAACCA	6900
AGCAAAACAA	GCAACCAAG	CAAACCAAGC	AAACCAAGCA	ATGCTAAAAA	ACAATTTATA	6960
TGATAAACTA	AAACATACTC	CATACCATGG	CAATACAAGG	GATTTAATAA	TATGACAAAA	7020
GAAAATTTAC	AAAGTGTTC	ACAAATACG	ACCGCTTCAC	TTGTAGAATC	AAACAACGAC	7080
CAAACTTCCC	TGCAATACT	TAAACAACCA	CCCAACCCA	ACCTATTACG	CCTGGAACAA	7140
CATGTCGCCA	AAAAAGATTA	TGAGCTTGCT	TGCCGCGAAT	TAATGGCGAT	TTTGGAAAAA	7200
ATGGACGCTA	ATTTGGAGG	CGTTCACGAT	ATTGAATTTG	ACGCACCTGC	TCAGCTGGCA	7260
TATCTACCG	AAAAACTACT	AATTCATTTT	GCCACTCGTC	TCGCTAATGC	AATTACAACA	7320
CTCTTTTCCG	ACCCCGAATT	GGCAATTTCC	GAAGAAGGGG	CATTAAAGAT	GATTAGCCTG	7380
CAACGCTGGT	TGACGCTGAT	TTTGGCTCT	TCCCCCTACG	TTAACGCAGA	CCATATTCTC	7440
AATAAATATA	ATATCAACCC	AGATTCGAA	GGTGGCTTTC	ATTAGCAAC	AGACAACCTCT	7500
TCTATTGCTA	AATTCTGTAT	TTTTTACTTA	CCGGAATCCA	ATGTCAATAT	GAGTTTAGAT	7560
GCGTTATGGG	CAGGGAATCA	ACAACCTTGT	GCTTCATTGT	GTTTGCCTT	GCAGTCTTCA	7620

CGTTTATTG	GTACTGCATC	TGCGTTTCAT	AAAAGAGCGG	TGGTTTACA	GTGGTTTCCT	7680
AAAAAATCG	CCGAAATGCG	TAATTTAGAT	GAATTGCCTG	CAAATATCCT	TCATGATGTA	7740
TATATGCACT	GCAGTTATGA	TTTAGCAAAA	AACAAGCACG	ATGTTAAGCG	TCCATTAAAC	7800
GAAGTTGTC	GCAAGCATAT	CCTCAGCAG	GGATGGCAAG	ACCGCTACCT	TTACACCTTA	7860
GGTAAAAAG	ACGGCAAACC	TGTGATGATG	GTACTGCTTG	AACATTTTAA	TTCCGGGACAT	7920
TCGATTTATC	GCACGCATTC	AACCTCAATG	ATTGCTGCTC	GAGAAAAATT	CTAATTTAGTC	7980
GGCTTAGGCC	ATGAGGGCGT	TGATAACATA	GGTCGAGAAG	TGTTTGACGA	GTTCTTTGAA	8040
ATCAGTAGCA	ATAATATAAT	GGAGAGACTG	TTTTTTATCC	GTAACAGATG	CGAAACTTTTC	8100
CAACCCGACG	TGTTCTATAT	GCCAAGCATT	GGCATGGATA	TIACCACGAT	TTTTGTGAGC	8160
AACACTCGGC	TTGCCCTTAT	TCAAGCTGTA	GCCTTGGGTC	ATCCIGCCAC	TACGCATTCT	8220
GAATTTATTG	ATTATGTCAT	CGTAGAAGAT	GATTATGTGG	GCAGTGAAGA	TTGTTTATAGC	8280
GAAACCCCTT	TAGGCTTACC	CAAAGATGCC	CTACCTTATG	TACCATCTCG	ACTCGCCCCA	8340
CAAAAAGTGG	ATTATGTACT	CAGGGAAGAA	CCTGAAGTAG	TCAATATCGG	TATTGCCGCT	8400
ACCACAATGA	AATTAACCC	TGAATTTTTG	CTAACATGTC	AAGAAATCAG	AGATAAAGCT	8460
AAAGTCAAAA	TACATTTTCA	TTTCGCACTT	GGACAATCAA	CAGGCTTGAC	ACACCCCTTAT	8520
GTCAAATGGT	TTATCGAAG	CTATTTAGGT	GACGATGCCA	CTGCACATCC	CCACGCACCT	8580
TATCAGGATT	ATCTGGCAAT	ATTGCGTGAT	TGCGATATGC	TACTAAATCC	GTTTCTTTTC	8640
GGTAATACTA	ACGGCATAAT	TGATATGGTT	ACATTAGGTT	TAGTTGGTGT	ATGCAAAACG	8700
GGGGATGAAG	TACATGAACA	TATTGATGAA	GGTCTGTTTA	AACGCTTAGG	ACTACCAGAA	8760
TGGCTGATAG	CCGACACACG	AGAAACATAT	ATTGAATGTG	CTTTGCGTCT	AGCAGAAAAC	8820
CATCAAGAAC	GCCTTGAAGT	CGCTCGTTAC	ATCATAGAAA	ACAACGGCTT	ACAAAAGCTT	8880
TTTACAGCGC	ACCTCTGCTC	ATTGGGCAAA	ATACTGCTTA	AGAAAACAAA	TGAATGGAAG	8940
CGGAAGCACT	TGAGTAAAAA	ATAACGGTTT	TTTAAAGTAA	AAGTGGCGGT	AATTTTCAAA	9000
GCGTTTTAAA	AACCTCTCAA	AAATCAACCG	CACCTTTTATC	TTTATAACGC	TCCCGCGCGC	9060
TGACAGTTTA	TCCTTTCTT	AAAATACCCA	TAAATTTGTG	GCAATAGTTG	GGTAATCAAA	9120
TTCAATGTGT	GATACGGCAA	ACTAAAGACG	GCGCGTTCTT	CGGCAGTCAT	C	9171

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 9323 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CGCCACTTCA ATTTTGGATT GTTGAAATTC AACTAACCAA AAAGTGGCGT TAAATCTGT

60

GGAGAAAATA GGTGTAGTG AAGAACGAGG TAATGTGTTCA AAAGGATAAA GCTCTCTTAA	120
TTGGGCATTG GTTGGCGTTT CTTTTTCGGT TAATAGTAAA TTATATTCTG GACGACTATG	180
CAATCCACCA ACAACTTTAC CGTTGGTTTT AAGCGTTAAT GTAAGTCTCT GCTCTCTCTG	240
GCGAATACGT AATCCCATTT TTTGTTTAGC AAGAAAATGA TCGGATAAT CATAATAGGT	300
GTTGCCCAA AATAAATTTT GATGTTCTAA AATCATAAAT TTGCAAGAT ATTGTGGCAA	360
TTCAATACCT ATTTGTGGCG AAATCGCCAA TTTAATTCA ATTTCTTGTA GCATAATATT	420
TCCCACTCAA ATCAACTGCT TAAATATACA AGATAATAAA AATAAATCAA GATTTTTGTG	480
ATGACAAACA ACAATTACAA CACCTTTTTT GCAGTCTATA TGCAATATT TTAaaaaaat	540
AGTATAATC CGCCATATAA AATGATATAA TCCTTCATCT TTCATCTTC ATCTTCATC	600
TTTCATCTTT CATCTTCATT CTTTCACTTT TCATCTTTCA TCTTCATCT TTCATCTTTC	660
ATCTTTCATC TTTTCATCTT CACATGAAAT GATGAACCGA GGAAGGGGAG GGAGGGGCAA	720
GAATGAAGAG GGAGCTGAAC GAACGCAAAAT GATAAAGTAA TTTAATTGTT CAACTAACCT	780
TAGGAGAAAA TATGAACAAG ATATATCGTC TCAAATTCAG CAAACGCCCTG AATGCTTTGG	840
TTGCTGTGTC TGAATTGGCA CGGGGTTGTG ACCATTCCAC AGAAAAAGGC AGCGAAAAAC	900
CTGCTCGCAT GAAAGTGCCT CACTTAGCGT TAAAGCCACT TTCCGCTATG TTACTATCTT	960
TAGGTGTAAC ATCTATTCCA CAATCTGTTT TAGCAAGCGG CAATTAAACA TCGACCAAAA	1020
TGAAATGGTG CAGTTTTTAC AAGAAAACAA GTAATAAAC CATTATCCGC AACAGTGTG	1080
ACGCTATCAT TAATTGGAAA CAATTTAACA TCGACCAAAA TGAATGGTG CAGTTTTTAC	1140
AAGAAAACAA CAACTCCGCC GTATTCAACC GTGTTACATC TAACCAAATC TCCCAATTAA	1200
AAGGGATTTT AGATTCTAAR GGACAAGTCT TTTTAATCAA CCCAAATGGT ATCACAATAG	1260
GTAAGACGC AATTATTAAC ACTAATGGCT TTACGGCTTC TACGCTAGAC ATTTCTAAGC	1320
AAAAATCAAA GCGCGTAAT TTCACCTTCG AGCAAAACCA AGATAAAGCG CTCGCTGAAA	1380
TTGTGAATCA CGGTTTAATT ACTGTCGGTA AAGACGGCAG TGTAAATCTT ATTGTGGCA	1440
AAGTGAAAAA CGAGGGTGTG ATTAGCGTAA ATGGTGGCAG CATTTCCTTA CTCGAGGGC	1500
AAAAATACAC CATCAGCGAT ATAATAAACC CAACCATTAAC TTACAGCATT GCGCGCCTG	1560
AAATGAAGC GGTCAATCTG GCGGATATTT TTGCCAAAGG CGGTAAACAT AATGTCCGTG	1620
CTGCCACTAT TCGAAACCAA GGTAAACTTT CTGCTGATTC TGTAAGCAAA GATAAAAGCG	1680
GCAATATTGT TCTTTCCGCC AAAGAGGGTG AAGCGGAAAT TGGCGGTGTA ATTTCCGCTC	1740
AAATCAGCA AGCTAAAGGC GGCAAGCTGA TGATAAAGTC CGATAAAGTC ACATTAaaaa	1800
CAGGTGAGT TATCGACCTT TCAGGTAAG AAGGGGGAGA AACTTACCTT GCGGTGACG	1860
AGCGGGGCGA AGGTAAAAAC GGCATTCAAT TAGCAAGAA AACCTCTTTA GAAAAAGGCT	1920
CAACCATCAA TGATCAGGC AAAGAAAAAG GCGGACGCGC TATTGTGTGG GCGATATTG	1980
CGTTAATTGA CGGCAATATT AACGCTCAAG GTAGTGTGTA TATCGCTAAA ACCGTGGTT	2040
TTGTGGAGAC ATCGGGGCAT TATTTATCCA TTGACAGCAA TGCAATTGTT AAAACAAAAG	2100

AGTGTTGCT AGACCCTGAT GATGTAACAA TTGAAGCCGA AGACCCCTT CGCAATAATA	2160
CCGGTATAAA TGATGAATTC CCAACAGGCA CCGGTGAAGC AAGCGACCTT AAAAAAATA	2220
GCGAACTCAA AACACGCTA ACCAATACAA CTATTTCAAA TTATCTGAAA AACGCTTGA	2280
CAATGAATAT AACGGCATCA AGAAAACCTA CCGTTAATAG CTCATCAAC ATCGGAAGCA	2340
ACTCCCACTT AATTCTCCAT AGTAAAGGTC AGCGTGGCGG AGGCGTTCAG ATTGATGGAG	2400
ATATTACTTC TAAAGGCGGA AATTAAACCA TTTATTCTGG CGGATGGGTT GATGTTTATA	2460
AAAATATTAC GCTTGATCAG GGTTTTTTAA ATATTACCGC CGCTCCGTA GCTTTTGAAG	2520
GTGGAATAA CAAAGCACGC GACGCGCAA ATGCTAAAAT TGTCGCCAG GGCATGTAA	2580
CCATTACAGG AGAGGAAAA GATTTCAGGG CTAACAACGT ATCTTAAAC GGAACGGGTA	2640
AAGGCTGAA TATCATTICA TCAGTGAATA ATTTAACCCA CAATCTTAGT GGCACAAATA	2700
ACATATCTGG GAATATAACA ATTAACCAA CTACGAGAAA GAACACCTCG TATTGCAAAA	2760
CCAGCCATGA TTCGCACTGG AACGTCAGTG CTCTTAATCT AGAGACAGGC GCAAAATTTA	2820
CCTTTATTAA ATACATTICA AGCAATAGCA AAGGCTTAAC AACACAGTAT AGAAGCTCTG	2880
CAGGGGTGAA TTTTAAAGGC GTAAATGGCA ACATGTCATT CAATCTCAA GAAGGAGCGA	2940
AAGTTAATTT CAAATTAATA CCAAACGAGA ACATGAACAC AAGCAAACCT TTACCAATTC	3000
GGTTTTAGC CAATATCACA GCCACTGGTG GGGGCTCTGT TTTTTTGAT ATATATGCCA	3060
ACCATCTCGG CAGAGGGGCT GAGTTAAAA TGAGTGAAT TAATATCTCT AACGCGGCTA	3120
ATTTTACCTT AAATCCCAT GTTCGCGGCG ATGACGCTTT TAAATCAAC AAAGACTTAA	3180
CCATAATGC AACCAATTCA AATTTCAGCC TCAGACAGAC GAAAGATGAT TTTTATGACG	3240
GGTACGCACG CAATGCCATC AATTCAACCT ACAACATATC CATTCTGGG GGTATATGCA	3300
CCCTGGTGG ACAAACTCA AGCAGCAGCA TTACGGGAA TATTACTATC GAGAAAGCAG	3360
CAAATGTAC GCTAGAAGCC AATAAGCCC CTAATCAGCA AAACATAAGG GATAGAGTTA	3420
TAAACTTGG CAGCTTGCTC GTTAATGGGA GTTTAAGTTT AACTGGCGAA ARTGCAGATA	3480
TTAAAGGCAA TCTCACTATT TCAGAAAGCG CCACCTTTTAA AGGAAAGACT AGAGATACCC	3540
TAAATATCAC CGGCAATTTT ACCAATAATG GCACCTGCCA AATTAATATA ACACAAGGAG	3600
TGTTAAACT TGGCAATGTT ACCAATGATG GTGATTTAAA CATTACCACT CACGCTAAAC	3660
GCAACCAAAG AAGCATCATC GCGGAGATA TAATCAACAA AAAAGGAAGC TTAATATTA	3720
CAGACAGTAA TAATGATGCT GAAATCCAAA TTGGCGGCAA TATCTCGCAA AAAGAAGGCA	3780
ACCTCAGAT TTCTCCGAT AAAATTAATA TCACCAAACA GATAACAATC AAAAGGGTA	3840
TTGATGGAGA GGACTCTAGT TCAGATGCGA CAAGTAATGC CAACCTAATC ATTTAAACCA	3900
AAGAATTGAA ATTGACAGAA GACCTAAGTA TTTACGGTTT CAATAAGCA GAGATTACAG	3960
CCAAAGATG TAGAGATTTA ACTATTGGCA ACAGTAATGA CGGTAACAGC GGTGCCGAAG	4020
CCAAACAGT AACTTTTAAAC AATGTTAAAG ATTCAAAAAT CTCTGCTGAC GGTCAATAG	4080
TGACACTAAA TAGCAAAGTG AAAACATCTA GCAGCAATG CGGACGTGAA AGCAATAGCG	4140

ACAACGATAC CGGCTTAACT ATTACTGCAA AAAATGTAGA AGTAAACAAA GATATTACTT	4200
CTCTCAAAAC AGTAAATATC ACCGCGTCGG AAAAGGTTAC CACCACAGCA GGCTCGACCA	4260
TTAACGCAAC AAATGGCAAA GCAAGTATTA CAACCAAAAC AGGTGATATC AGCGGTACGA	4320
TTTCCGGTAA CACGCTAAGT GTTAGCGCGA CTGGTGATT AACCACTAAA TCCGGCTCAA	4380
AAATTGAAGC GAAATCGGGT GAGGCTAATG TAACAAGTGC AACAGGTACA GTTGGCGGTA	4440
CAATTTCCGG TAATACGGTA AATGTTACGG CAAACGCTGG CGATTTAACA GTTGGGAATG	4500
GCGCAGAAAT TAATGCGACA GAAGGAGCTG CAACCTTAAC CGCAACAGGG AATACCTTGA	4560
CTACTGAAGC CGGTCTTAGC ATCACTTCAA CTAAGGGTCA GGTAGACCTC TTGGCTCAGA	4620
ATGGTAGCAT CGCAGGAAGC ATTAATGTCTG CTAATGTGAC ATTAAATACT ACAGGCACCT	4680
TAACCACCGT GGCAGGCTCG GATATTAAGC CAACCAGCGG CACCTTGGTT ATTAACGCAA	4740
AAGATGCTAA GCTAAATGGT GATGCATCAG GTGATAGTAC AGAAGTGAAT CGAGTCAACG	4800
ACTGGGGATT TGGTAGTGTG ACTGCGGCAA CCTCAAGCAG TGTGAATATC ACTGGGGATT	4860
TAAACACAGT AAATGGGTAA AATATCATT CGAAAGATGG TAGAAACACT GTGCGCTTAA	4920
GAGGCAAGGA AATTGAGGTG AAATATATCC AGCCAGGTGT AGCAAGTGTA GAAGAAGTAA	4980
TTGAAGCGAA ACGCGTCTCT GAAAAAGTAA AAGATTTATC TGATGAAGAA AGAGAAACAT	5040
TAGCTAAACT TGGTGTAAGT GCTGTACGTT TTGTTGAGCC AAATAATACA ATTACAGTCA	5100
ATACACAAAA TGAATTTACA ACCAGACCGT CAAGTCAAGT GATAATTTCT GAAGGTAAGG	5160
CGTGTCTCTC AAGTGGTAAT GCGCAGCAGG TATGTACCAA TGTGCTGAC GATGGACAGC	5220
CGTAGTCAGT AATTGACAAG GTAGATTCA TCCTGCAATG AAGTCATTTT ATTTTCGTAT	5280
TATTTACTGT GTGGGTAAAG GTTCAGTACG GGCTTTACCC ATCTTGTAAG AAATTACGGA	5340
GAATACAATA AAGTATTTT AACAGGTTAT TATTATGAAA AATATAAAAA CGAGATTAAA	5400
ACTCAGTGCA ATATCAGTAT TGCTTGGCCT GGCTTCTTCA TCATTGTATG CAGAAGAAGC	5460
GTTTTAGTA AAAGGCTTTC AGTTATCTGG TGCACCTGAA ACTTTAAGTG AAGACGCCCA	5520
ACTGTCTGTA GCAAAATCTT TATCTAAATA CCAAGGCTCG CAACTTTAA CAAACCTAAA	5580
AACAGCACAG CTTGAATTAC AGGCTGTGCT AGATAAGATT GAGCCAAATA AATTTGATGT	5640
GATATTGCCG CAACAAACCA TTACGGATGG CAATATCATG TTTGAGCTAG TCTCGAAATC	5700
AGCCGCGAAA AGCCAAAGTT TTTATAAGGC GAGCCAGGGT TATAGTGAAG AAAATATCGC	5760
TCGTAGCCTG CCATCTTTGA AACCAAGGAA AGTGTATGAA GATGGTCGT AGTGGTTGCA	5820
TTTGCGTGAA TTTAATATGG CAAAAGAAAA CCCGCTTAAG GTTACCCGTG TACATTACGA	5880
ACTAAACCTT AAAAACAAAA CCTCTAATT GATAATTGCG GGCTTCTCGC CTTTGTGTAA	5940
AACGCGTAGC TTTATTCTCT ATGATAATT CCGCGCGAGA GAGTTTAACT ACCAACGTGT	6000
AAGCTTGGGT TTTGTTAATG CCAATTTAAC TGGTCATGAT GATGTGTTAA TTATACCAGT	6060
ATGAGTTATG CTGATTCTAA TGATATCGAC GGCTTACCAA GTGCGATTAA TCGTAAATTA	6120
TCAAAGGTC AATCTATCTC TGCGAATCTG AAATGGAGTT ATTATCTCCC AACATTTAAC	6180

CTTGGCATGG AAGACCAATT TAAATTAAT	TTAGGTACA ACTACCGCCA TATTAATCAA	6240
AOCTCCGCGT TAAATCGCTT GGGTGAAACG	AAGAAAAAAT TTGCAGTATC AGSCGTAAGT	6300
GCAGGCATTG ATGGACATAT CCAATTTACC	CCTAAAAACAA TCTTTAATAT TGATTTAACT	6360
CATCATTTATT ACGCGAGTAA ATTACCAGGC	TCTTTTGGAA TGGAGCGCAT TGGCGAAACA	6420
TTTAATCGCA GCTATCACA TAGCACAGCC	AGTTTAGGGT TGAGTCAAGA GTTTGCTCAA	6480
GGTTGGCATT TTAGCAGTCA ATTATCAGGT	CAATTTACTC TACAAGATAT TAGCAGTATA	6540
GATTTATTCT CTGTAACAGG TACTTATGGC	GTACAGAGGT TTAATATACGG CGGTGCAAGT	6600
GGTGAGCGCG GTCTTGATG GCGTAATGAA	TTAAGTATGC CAAAATACAC CCGCTTCCAA	6660
ATCAGCCCTT ATGCGTTTTA TGATGCAGGT	CAGTTCCGTT ATAATAGCGA AAATGCTAAA	6720
ACTTACGGCG AAGATATGCA CACGGTATCC	TCTGCGGGTT TAGGCATTAA AACCTCTCCT	6780
ACACAAAAT TAAGCCTAGA TGCTTTTGTT	GCTCGTCGCT TTGCAAAATGC CAATAGTGAC	6840
AATTTGAATG GCAACAAAAA ACGCACAAAG	TCACCTACAA CCTTCTGGGG GAGATTAACA	6900
TTCAGTTTCT AACCTGAAA TTTAATCAAC	TGGTAAGCGT TCCGCCTACC AGTTTATAAC	6960
TATATGCTTT ACCCGCCAAT TTACAGTCTA	TAGGCAACCC TGTTTTTACC CTTATATATC	7020
AAATAACCAA GCTAAGCTGA GCTAAGCAAA	CCAAGCAAAC TCAAGCAAGC CAAGTAATAC	7080
TAAAAAACAA ATTTATATGA TAAACTAAG	TATACTCCAT GCATGGCGA TACAAGGGAT	7140
TTAATAATAT GACAAAAGAA AATTTGCAA	ACGCTCCTCA AGATGCGACC GCTTTACTTG	7200
CGGAATTAAG CAACAATCAA ACTCCCTGCG	GAATATTTAA ACAACCACGC AAGCCGAGCC	7260
TATTACGCTT GGAACAACAT ATCGCAAAAA	AAGATTATGA GTTTGCTTGT CGTGAATTAA	7320
TGSGTATTCT GGAAAAATG GACGCTAATT	TTGGAGCGCT TCACGATATT GAATTTGACG	7380
CACCCGCTCA GCTGGCATAT CTACCCGAAA	AATTACTAAT TTATTTTGCC ACTCGTCTCG	7440
CTAATGCAAT TACAACACTC TTTTCCGACC	CCGAATTTGC AATTTCTGAA GAAGGGGCGT	7500
TAAAGATGAT TAGCCTGCAA CGCTGGTTGA	CGCTGATTTT TGCTCTCTCC CACTACGTTA	7560
ACGCAGACCA TATTCTCAAT AAATATAATA	TCAACCCAGA TTCCGAAGGT GGCCTTCATT	7620
TAGCAACAGA CAACTCTTCT ATTGCTAAAT	TCTGTATTTT TTACTTACCC GAATCCAATG	7680
TCAATATGAG TTTAGATGCG TTATGGGCAG	GGAATCAACA ACTTTGTGCT TCATTGTGTT	7740
TTGCGTTGCA GTCTTCACGT TTTATTGGTA	CCGCATCTGC GTTTCATAAA AGAGCGGTGG	7800
TTTTACAGTG GTTTCCTAAA AAACCTGCGG	AAATTGCTAA TTTAGATGAA TTGCTGCAA	7860
ATATCCTTCA TGATGTATAT ATGCACTGCA	GTTATGATTT AGCAAAAAAC AAGCAGATG	7920
TTAAGCGTCT ATTAACGAA CTTGTCCGCA	AGCATATCCT CACGCAAGGA TGGCAAGACC	7980
GCTACCTTTA CACCTTAGGT AAAAAGGACG	GCAAACCTGT GATGATGGTA CTGCTTGAAC	8040
ATTTTAATTC GGGACATTG ATTTATCGTA	CACATTCAC TTCAATGATT GCTGCTCGAG	8100
AAAAATTTCTA TTTAGTCGGC TTAGGCCATG	AGGGCGTTGA TAAATAGGT CGAGAAGTGT	8160
TTGACGAGTT CTTTGAAATC AGTAGCAATA	ATATAATGGA GAGACTGTTT TTTATCCGTA	8220

AACAGTGC GA AACTTTCCAA CCCGCAGTGT TCTATATGCC AAGCATTGGC ATGGATATTA	8280
CCACGATTTT TGTGAGCAAC ACTCGGCTTG CCCCTATTCA AGCTGTAGCC CTGGGTCATC	8340
CTGCCACTAC GCATTCGTAA TTTATTGATT ATGTCATCGT AGAAGATGAT TATGTGGGCA	8400
GTGAAGATTG TTTCAGCGAA ACCCTTTTAC GCTTACCCAA AGATGCCCTA CCTTATGTAC	8460
CTTCTGCACT CGCCCCCAA AAAGTGGATT ATGTACTCAG GGAAAACCTT GAAGTAGTCA	8520
ATATCGGTAT TGCCGCTACC ACAATGAAAT TAAACCCTGA ATTTTTCCTA ACATTGCAAG	8580
AAATCAGAGA TAAAGCTAAA GTCAAAATAC ATTTTCATTT CGCACTTGGA CAATCAACAG	8640
GCTTGACACA CCTTATTGTC AAATGGTTTA TCGAAAAGCTA TTTAGGTGAC GATGCCACTG	8700
CACATCCCCA CGCACCTTAT CACGATTATC TGGCAATATT GCGTGATTGC GATATGCTAC	8760
TAAATCCGTT TCCTTTCGGT AATACTAACG GCATAATTGA TATGGTTACA TTAGGTTTAG	8820
TTGGTGTATG CAAAACGGGG GATGAAGTAC ATGAACATAT TGATGAAGGT CTGTTTAAAC	8880
GCTTAGGACT ACCAGAATGG CTGATAGCCG ACACACGAGA AACATATATT GAATGTGCTT	8940
TGCGCTTAGC AGAAAACCAT CAAGAACGCC TTGAACCTCG TCGTTACATC ATAGAAAACA	9000
ACGGCTTACA AAAGCTTTTT ACAGGCGACC CTCGTCCATT GGGCAAAATA CTGCTTAAAG	9060
AAACAAATGA ATGGAAGCGG AAGCACTTGA GTAAAAAATA ACGGTTTTTT AAAGTAAAG	9120
TGCGGTTAAT TTTCAAACGG TTTTAAAAAC CTCTCAAAAA TCAACCGCAC TTTTATCTTT	9180
ATAACGATCC CGCACGTGA CAGTTTATCA GCCTCCCGCC ATAAACTCC GCCTTTCATG	9240
GCGGAGATTT TAGCCAAAAC TGGCAGAAAT TAAAGGCTAA AATCACCAAA TTGCACCACA	9300
AAATCACCAA TACCACAAA AAA	9323

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4794 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATGAACAAGA TATATCGTCT CAAATTCAGC AAACGCCTGA ATGCTTTGGT TGCTGTGTCT	60
GAATTGACAC GGGGTTGTGA CCATTCCACA GAAAAAGGCA GTGAAAAACC TGTTCTGTAG	120
AAAGTAGGCC ACTTGCGGTT AAAGCCACTT TCGCTATAT TGCTATCTTT GGGCATGGCA	180
TCCATTCCGC AATCTGTTTT AGCGAGCGGT TTACAGGGAA TGAGCGTCGT ACACGGTACA	240
GCAACCATGC AAGTAGACGG CAATAAAACC ACTATCCGTA ATAGCGTCAA TGCTATCATC	300
AATTGGAAC AATTTAACAT TGACCAAAAT GAAATGGTGC AGTTTTTACA AGAAAGCAGC	360
AACTCTGCCG TTTTCAACGG TGTTACATCT GACCAAATCT CCCAATTAAA AGGGATTTTA	420

GATTCTAACG GACAAGTCTT TTTAATCAAC CCAATGTA TCACAATAGG TAAAGACGCA	480
ATTATTAAACA CTAATGGCCT TACTGCTTCT ACGCTAGACA TTCTAACGA AAACATCAAG	540
CGCGGTAATT TCACCCCTGA GCAAACCAAG GATAAAGCAC TCGCTGAAAT CGTGAATCAC	600
GGTTTAATTA CCGTTGGTAA AGACGGTAGC GTAACACCTTA TTGGTGGCAA AGTGAAAAAC	660
GAGGGCGTGA TTAGCGTAAA TGGCGGTAGT ATTTCTTAC TTGCAGGGCA AAAAATCACC	720
ATCAGCGATA TAATAAATCC AACCATCACT TACAGCATTG CTGCACCTGA AAACGAAGCG	780
ATCAATCTGG GCGATATTTT TGCCAAAGGT GGTAAACATTA ATGTCCGCGC TGCCACTATT	840
CGCAATAAAG GTAAACTTTC TGCCGACTCT GTAAGCAAAG ATAAAAGTGG TAACATTGTT	900
CTCTCTGCCA AAGAAGGTGA AGCGGAAATT GCGCGTGTAA TTTCGCTCA AAATCAGCAA	960
GCCAAAGGTG GTAAGTTGAT GATTACAGGC GATAAAGTTA CATTGAAAAA GGGTGCAGTT	1020
ATCGACCTTT CGGGTAAAGA AGGGGGAGAA ACTTATCTTG GCGGTGACGA GCGTGGCGAA	1080
GGTAAAAACG GCATTCAATT AGCAAAGAAA ACCACTTTAG AAAAAGGCTC AACAAATTAAT	1140
GTGTCAAGTA AAGAAAAAGG TGGGCGCGCT ATTGTATGGG GCGATATTGC GTTAATTGAC	1200
GGCAATATTA ATGCCCAAGG TAAAGATATC GCTAAAAGT GTGGTTTTGT GGAGACGTCG	1260
GGCATTACTT TATCCATTGA TGATAACGCA ATTGTTAAAA CAAAAGAATG GCTACTAGAC	1320
CCAGAGAATG TGACTATTGA AGCTCCTTCC GCTTCTCGCG TCGAGCTGGG TGCCGATAGG	1380
AATTCCCACT CGGCAGAGGT GATAAAGTG ACCCTAAAAA AAAATAACAC CTCCTTGACA	1440
ACACTAACCA ATACAACCAT TTCAAATCTT CTGAAAAAGT CCCAGCTGGT GAACATAACG	1500
GCAAGGAGAA AACTTACCGT TAATAGCTCT ATCAGTATAG AAGAGGGCTC CCACCTTAAT	1560
CTCCACAGTG AAGGTCAGGG CGGTCAAGGT GTTCAGATTG ATAAAGATAT TACTTCTGAA	1620
GGCGGAAATT TAACCATTTA TTCTGGCGGA TGGGTTGATG TTCATAAAAA TATTACGCTT	1680
GGTAGCGGCT TTTTAAACAT CACAACATAA GAAGGAGATA TCGCCTTGA AGACAAGCTT	1740
GGACGGAACA ACCTAACCAT TACAGCCCAA GGGACCATCA CCTCAGGTAA TAGTAAACGC	1800
TTTAGATTTA ACAACGCTCT TCTAAACAGC CTGGCGGAA AGCTGAGCTT TACTGACAGC	1860
AGAGAGGACA GAGGTAGAAG AACTAAGGGT AATATCTCAA ACAAAATTGA CGGAACGTTA	1920
AACATTTCCG GAAGTGTAGA TATCTCAATG AAAGCACCCA AAGTCAGCTG GTTTTACAGA	1980
GACAAAGGAC GCACCTACTG GAACGTAACC ACITTTAAATG TTACCTCGGG TAGTAAATTT	2040
AACCTCTCCA TTGACAGCAC AGGAAGTGGC TCAACAGGTC CAAGCATACG CAATGCAGAA	2100
TTAAATGGCA TAACATTAA TAAAGCCACT TTTAATATCG CACAAGGCTC AACAGCTAAC	2160
TTTAGCATCA AGGCATCAAT AATGCCCTTT AAGAGTAAAG CTAACCTACGC ATTATTTAAT	2220
GAAGATATTT CAGTCTCAGG GGGGGGTAGC CTTAATTICA AACTTAACGC CTCATCTAGC	2280
AACATACAAA CCCCTGGCGT AATTATAAAA TCTCAAAACT TTAATGTCTC AGGAGGGTCA	2340
ACTTTAAATC TCAAGGCTGA AGGTTCAACA GAAACCGCTT TTTCATAGA AAATGATTTA	2400
AACCTAAACG CCACCGGTGG CAATATAACA ATCAGACAAG TCGAGGGTAC CGATTACGC	2460

GTCAACAAAG GTGTGCGAGC CAAAAAAAC ATAACCTTTTA AAGGGGGTAA TATCACCTTC	2520
GGCTCTCAA AAGGCACAAC AGAAATCAAA GGCAATGTTA CCATCAATAA AAACACTAAC	2580
GCTACTCTTT GTGTGCGGAA TTTTGCCGAA AACAAATGCG CTTTAAATAT AGCAGGAAAT	2640
GTTATTAATA ATGGCAACCT TACCACTGCC GGCTCCATTA TCAATATAGC CGGAAATCTT	2700
ACTGTTTCAA AAGGCGCTAA CCTTCAAGCT ATAACAAAT ACACCTTTAA TGTAGCCGGC	2760
TCATTTGACA ACAATGGCGC TTCAAACATT TCCATTGCCA GAGGAGGGGC TAAATTTAAA	2820
GATATCAATA ACACCAAGTAG CTTAAATATT ACCACCAACT CTGATACCAC TTACCGCACC	2880
ATTATAAAG GCAATATATC CAACAAATCA GGTGATTGTA ATATTATTGA TAAAAAAGC	2940
GACGCTGAAA TCCAAATTGG CGGCAATATC TCACAAAAAG AAGGCAATCT CACAATTCT	3000
TCTGATAAAG TAAATATTAC CAATCAGATA ACAATCAAAG CAGGCGTTGA AGGGGGGCGT	3060
TCTGATTCAA GTGAGGCAGA AAATGCTAAC CTAACATTC AAACCAAAGA GTTAAATTTG	3120
GCAGGAGACC TAAATATTTC AGGCTTTAAT AAAGCAGAAA TTACAGCTAA AAATGGCAGT	3180
GATTTAACTA TTGGCAATGC TAGCGTGGT AATGCTGATG CTA AAAAAGT GACTTTTGAC	3240
AAGGTTAAAG ATTCAAAAAT CTCGACTGAC GGTCACAATG TAACACTAAA TAGCGAAGTG	3300
AAAACGCTCTA ATGGTAGTAG CAATGCTGGT AATGATAACA GCACCGGTTT AACCATTTCC	3360
GCAAAAGATG TAACGGTAAA CAATAACGTT ACCTCCACAC AGACAATAAA TATCTCTGCC	3420
GCAGCAGGAA ATGTAACAAC CAAAGAAGGC ACAACTATCA ATGCAACCAC AGGCAGCGTG	3480
GAAGTAACTG CTCAAAATGG TACAATTAAA GGCAACATTA CCTGCAAAA TGTAAACAGTG	3540
ACAGCAACAG AAAATCTTGT TACCACAGAG AATGCTGTCA TTAATGCAAC CAGCGGCACA	3600
GTAACATTA GTACAAAAAC AGGGGATATT AAAGGTGGAA TTGAATCAAC TTCCGGTAAT	3660
GTAATATTA CAGCGAGCGG CAATACACTT AAGGTAAGTA ATATCACTGG TCAAGATGTA	3720
ACAGTAACAG CGGATGCGAG AGCCTTGACA ACTACAGCAG GCTCAACCAT TAGTGCGACA	3780
ACAGGCAATG CAAATATTAC AACCAAAACA GGTGATATCA ACGGTAAGT TGAATCCAGC	3840
TCCGCTCTG TAACACTTGT TGCAACTGGA GCAACTCTTG CTGTAGGTAA TATTTAGGT	3900
AACACTGTTA CTATTACTGC SGATAGCGGT AAATTAACCT CCACAGTAGG TTCTACAATT	3960
AATGGGACTA ATAGTGTAA CACCTCAAGC CAATCAGGCG ATATTGAAG TACAATTCT	4020
GGTAATACAG TAAATGTTAC AGCAAGCACT GGTGATTAA CTATTGSAAT TAGTGCAAAA	4080
GTTGAAGCGA AAAATGGAGC TGCAACCTTA ACTGCTGAAT CAGGCAAAAT AACCAACCAA	4140
ACAGGCTCTA GCATTACCTC AAGCAATGGT CAGACAACTC TTACAGCCAA GGATAGCAGT	4200
ATCGCAGGAA ACATTAATGC TGCTAATGTG ACGTTAAATA CCACAGGCAC TTAACTACT	4260
ACAGGGGATT CAAAGATTA CGCAACCACT GGTACCTTAA CAATCAATGC AAAAGATGCC	4320
AAATTAGATG GTGCTGCATC AGGTGACCGC ACAGTAGTAA ATGCAACTAA CGCAAGTGGC	4380
TCTGTTAAG TGACTGCGAA AACCTCAAGC AGCGTGAATA TCACCGGGA TTTAAACACA	4440
ATAATGGGT TAAATATCAT TTCGGAATG GGTAGAAACA CTGTGCGCTT AAGAGGCAAG	4500

GAAATTGATG TGAATATAT CCAACCAGGT GTAGCAAGCG TAGAAGAGGT AATTGAAGCG	4560
AAACGCGTCC TTGAGAAGGT AAAAGATTTA TCTGATGAAG AAAGAGAAAC ACTAGCCAAA	4620
CTTGCTGTAA GTGCTGTACG TTCTGTTAG CCAAATAATG CCATTACGGT TAATACACAA	4680
AACGAGITTA CAACCAAAACC ATCAAGTCAA GTGACAAATT CTGAAGGTAA GCGGTGTTTC	4740
TCAAGTGGTA ATGGCGCACG AGTATGTACC AATGTTGCTG ACGATGGACA GCAG	4794

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4803 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ATGAACAAGA TATATCGTCT CAAATTTCAGC AAACGCTCGA ATGCTTTGGT TGCTGTGTCT	60
GAATTGACAC GGGGTTGTGA CCATTCCACA GAAAAAGGCA GTGAAAAACC TGTTCTGTACG	120
AAAGTACGCC ACTTGGCGTT AAAGCCACTT TCCGCTATAT TGCTATCTTT GGGCATGGCA	180
TCCATTCCGC AATCTGTTTT AGCGAGCGGT TTACAGGGAA TGAGCGTCGT ACACGTTACA	240
GCAACCATGC AAGTAGACGG CAATAAAACC ACTATCCGTA ATAGCGTCAA TGCTATCATC	300
AATTGGAAC AATTTAACAT TGACCAAAAT GAAATGGTGC AGTTTTTACA AGAAAGCAGC	360
AATCTGCCG TTTTCAACCG TGTTACATCT GACCAAATCT CCCAATTTAA AGGGATTTTA	420
GATTCTAACG GACAAGTCTT TTTAATCAAC CCAATGGTA TCACAATAGG TAAAGACGCA	480
ATTATTAACA CTAATGGCTT TACTGCTTCT ACGCTAGACA TTTCTAACGA AAACATCAAG	540
GCGCGTAATT TCACCCCTGA GCAAACCAAG GATAAAGCAC TCGCTGAAAT CGTGAATCAC	600
GGTTTAATTA CCGTTGTGTA AGACGGTAGC GTAAACCTTA TTGGTGGCAA AGTGAAGAAC	660
GAGGGCGTGA TTAGCGTAA TGGCGGTAGT ATTTCTTTAC TTGCAGGGCA AAAAATCACC	720
ATCAGCGATA TAATAAATCC AACCATCACT TACAGCATTG CTGCACCTGA AAACGAAGCG	780
ATCAATCTGG GCGATATTTT TGCCAAAGGT GGTAACATTA ATGTCCGCGC TGCCACTATT	840
CGCAATAAAG GTAAACTTTC TGCCGACTCT GTAAGCAAAG ATAAAGTGG TAACATTGTT	900
CTCTCTGCCA AAGAAGGTGA AGCGGAAATT GCGGTGTAA TTTCCGCTCA AAATCAGCAA	960
GCCAAAGGTG GTAAGTTGAT GATTACAGGT GATAAAGTCA CATTAAAAAC AGGTGCAGTT	1020
ATCGACCTTT CAGGTAAAGA AGGGGGAGAG ACTTATCTTG GCGGTGATGA GCGTGGCGAA	1080
GGTAAAAATG GTATTCAATT AGCGAAGAAA ACCTCTTTAG AAAAAGGCTC GACAATTAAT	1140
GTATCAGGCA AAAAAAAGG CGGGCGCGCT ATTGTATGGG GCGATATTGC ATTAATTAAT	1200
GGTAACATTA ATGCTCAAGG TAGCGATATT GCTAAAAC TGCGCTTTGT GGAACATCA	1260

GGACATGACT TATCCATTGG TGATGATGTG ATTGTTGACG CTAAAGAGTG GTTATTAGAC	1320
CCAGATGATG TGTCATTGA AACTCTTACA TCTGGACGCA ATAATACCGG CGAAAAACCA	1380
GGATATACAA CAGGAGATGG GACTAAAGAG TCACCTAAAG GTAATAGTAT TTCTAAACCT	1440
ACATTAAACA ACTCAACTCT TGAGCAAATC CTAAGAAGAG GTTCTTATGT TAATATCACT	1500
GCTAATAATA GAATTTATGT TAATAGCTCC ATCAACTTAT CTAATGGCAG TTTAACACTT	1560
CACACTAAAC GAGATGGAGT TAAATTAAC GGTGATATTA CCTCAAACGA AAATGGTAAT	1620
TTAACCATTA AAGCAGGCTC TTGGGTTGAT GTTCATAAAA ACATCACGCT TGGTACGGGT	1680
TTTTGAATA TTGTCGCTGG GGATTCTGTA GCTTTTGAGA GAGAGGGCGA TAAAGCACGT	1740
AACGCAACAG ATGCTCAAAT TACCGCACAA GGGACGATAA CCGTCAATAA AGATGATAAA	1800
CAATTTAGAT TCAATAATGT ATCTATTAAC GGGACGGGCA AGGGTTTAAA GTTTATTGCA	1860
AATCAAAATA ATTCTACTCA TAAATTTGAT GCGGAAATTA ACATATCTGG AATAGTAACA	1920
ATTAAACAAA CCACGAAAAA AGATGTTAAA TACTGGAATG CATCAAAAGA CTCTTACTGG	1980
AATGTTTCTT CTCCTACTTT GAATACGGTG CAAAAATTA CCTTTATAAA ATTCTGTGAT	2040
AGCGGCTCAA ATTCCCAAGA TTTGAGGTCA TCACGTAGAA GTTTGCGAG CGTACATTTT	2100
AACGGCATCG GAGGCAAAAC AAACCTCAAC ATCGGAGCTA ACGCAAAAGC CTTATTTAAA	2160
TTAAACCAAA ACGCGCTAC AGACCCAAAA AAAGAATTAC CTATTACTTT TAACGCCAAC	2220
ATTACAGCTA CCGGTAACAG TGATAGCTCT GTGATGTTTG ACATACAGCG CAATCTTACC	2280
TCTAGAGCTG CCGGCATAAA CATGGATTCA ATTAACATTA CCGCGGGCT TGACTTTTCC	2340
ATAACATCCC ATAATCGCAA TAGTAATGCT TTTGAAATCA AAAAAAGACTT AACTATAAAT	2400
GCAACTGGCT CGAATTTTAG TCTTAAGCAA ACGAAAGATT CTTTTATAA TGAATACAGC	2460
AAACAGGCCA TTAACCAAG TCATAATCTA ACCATTCTTG GCGGCAATGT CACTCTAGGT	2520
GGGGAATTT CAAGCAGTAG CATTACGGGC AATATCAATA TCACCAATAA AGCAATGTTT	2580
ACATTACAAG CTGACACCAG CAACAGCAAC ACAGGCTTGA AGAAAAGAAC TCTAATCTT	2640
GGCAATATAT CTGTTGAGG GAATTTAAGC CTAATGCTG CAAATGCAAA CATTGTCGGC	2700
AATCTTTCTA TTGAGAAGA TTCCACATTT AAAGGAGAAG CCAAGTACAA CCTAAACATC	2760
ACCGGCACCT TTACCAACAA CGGTACCGCC AACATTAATA TAAACAAGG AGTGGTAAAA	2820
CTCCAAGGCG ATATTATCA TAAAGGTGGT TTAATATCA CTACTAACGC CTCAGGCACT	2880
CAAAAAACCA TTATTAACGG AAATATAACT AACGAAAAAG GCGACTTAAA CATCAAGAAT	2940
ATTAAAGCGG ACGCGGAAT CCAATTGGC GGCAATATCT CAAAAAGA AGGCAATCTC	3000
ACAATTTCTT CTGATAAAGT AAATATTACC AATCAGATAA CAATCAAAGC AGGCGTTGAA	3060
GGGGGCGGTT CTGATTCAAG TGAGGCAGAA AATGCTAACC TAATATTCA AACCAAGAG	3120
TTAAATTTGG CAGGAGACCT AAATATTCA GGCTTTAATA AAGCAGAAAT TACAGCTAAA	3180
AATGGCAGTG ATTTAACAT TGCAATGCT AGCGGTGGTA ATGCTGATGC TAAAAAGTG	3240
ACTTTTGACA AGGTTAAGA TTCAAAAATC TCGACTGACG GTCACAATGT AACACTAAAT	3300

AGCGAAGTGA AAACGTCTAA TGGTAGTAGC AATGCGGTA ATGATAACAG CACCGGTTTA	3360
ACCATTTCGC CAAAAGATGT AACGGTAAAC AATAAGTTA CCTCCACAA GACAATAAAT	3420
ATCTCTGCGC CAGCAGGAAA TGTAACAACC AAAGAAGGTA CAACTATCAA TGCAACCACA	3480
GGCAGCGTGG AAGTAAGTGC TCAAAATGGT ACAAATAAG GCAACATTAC CTCGCAAAAT	3540
GTAACAGTGA CAGCAACAGA AAATCTTGTT ACCACAGAGA ATGCTGTGAT TAATGCAACC	3600
AGCGGCACAG TAAACATTAG TACAAAAACA GGGGATATTA AAGGTGGAAT TGAATCAACT	3660
TCCGGTAATG TAAATATTAC AGCGAGCGGC AATACACTTA AGGTAAGTAA TATCACTGGT	3720
CAAGATGTAA CAGTAACAGC GGATGCAGGA GCCTTGACAA CTACAGCAGG CTCAACCATT	3780
AGTGCACAA CAGGCAATGC AAATATTACA ACCAAAACAG GTGATATCAA CGGTAAAGTT	3840
GAATCCAGCT CCGGCTCTGT AACACTTGTT GCAACTGGAG CAACTCTTGC TGTAGGTAAT	3900
ATTCAGGTA ACACTGTTAC TATTACTGCG GATAGCGGTA AATTAACCTC CACAGTAGGT	3960
TCTACAATTA ATGGAGCTAA TAGTGTAACC ACCTCAAGCC AATCAGGCGA TATTGAAGGT	4020
ACAATTTCTG GTAATACAGT AAATGTTACA GCAAGCACTG GTGATTTAAC TATTGGAAT	4080
AGTGCAAAAG TTGAAGCGAA AAATGGAGCT GCAACCTTAA CTGCTGAATC AGGCAATTA	4140
ACCACCCAAA CAGGCTCTAG CATTACCTCA AGCAATGGTC AGACAACCTC TACAGCCAAG	4200
GATAGCAGTA TCGCAGGAAA CATTAAATGCT GCTAATGTGA CGTTAAATAC CACAGGCACT	4260
TTAATACTA CAGGGGATTC AAAGATTAAC GCAACCACTG GTACCTTAAC AATCAATGCA	4320
AAAGATGCCA AATTAGATGG TGCTGCATCA GGTGACCGCA CAGTAGTAAA TGCAACTAAC	4380
GCAAGTGGCT CTGGTAACGT GACTGCGAAA ACCTCAAGCA GCGTGAATAT CACCGGGGAT	4440
TTAAACACAA TAAATGGGTT AAATATCATT TCGGAAATG GTAGAAACAC TGTGCGCTTA	4500
AGAGGCAAGG AAATTGATGT GAAATATATC CAACCAGGTG TAGCAAGCGT AGAAGAGGTA	4560
ATTGAAGCGA AACGCGCTCT TGAGAAGGTA AAAGATTAT CTGATGAAGA AAGAGAAACA	4620
CTAGCCAAC CTGGTGTAAG TGCTGTACGT TTCGTTGAGC CAAATAATGC CATTACGGTT	4680
AATACACAAA ACGAGTTTAC AACCACCACT TCAAGTCAAG TGACAATTC TGAAGGTAAG	4740
GCGTGTCTCT CAAGTGGTAA TGGCGCACGA GTATGTACCA ATGTTGCTGA CGATGGACAG	4800
CAG	4803

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1599 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu
 1 5 10 15

Val Ala Val Ser Glu Leu Thr Arg Gly Cys Asp His Ser Thr Glu Lys
 20 25 30

Gly Ser Glu Lys Pro Val Arg Thr Lys Val Arg His Leu Ala Leu Lys
 35 40 45

Pro Leu Ser Ala Ile Leu Leu Ser Leu Gly Met Ala Ser Ile Pro Gln
 50 55 60

Ser Val Leu Ala Ser Gly Leu Gln Gly Met Ser Val Val His Gly Thr
 65 70 75 80

Ala Thr Met Gln Val Asp Gly Asn Lys Thr Thr Ile Arg Asn Ser Val
 85 90 95

Asn Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met
 100 105 110

Glu Gln Phe Leu Gln Glu Ser Ser Asn Ser Ala Val Phe Asn Arg Val
 115 120 125

Thr Ser Asp Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly
 130 135 140

Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
 145 150 155 160

Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
 165 170 175

Glu Asn Ile Lys Ala Arg Asn Phe Thr Leu Glu Gln Thr Lys Asp Lys
 180 185 190

Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
 195 200 205

Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
 210 215 220

Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
 225 230 235 240

Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
 245 250 255

Glu Asn Glu Ala Ile Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn
 260 265 270

Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Lys Gly Lys Leu Ser Ala
 275 280 285

Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
 290 295 300

Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
 305 310 315 320

Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
 325 330 335

Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
 340 345 350
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala
 355 360 365
 Lys Lys Thr Thr Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
 370 375 380
 Glu Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
 385 390 395 400
 Gly Asn Ile Asn Ala Gln Gly Lys Asp Ile Ala Lys Thr Gly Gly Phe
 405 410 415
 Val Glu Thr Ser Gly His Tyr Leu Ser Ile Asp Asp Asn Ala Ile Val
 420 425 430
 Lys Thr Lys Glu Trp Leu Leu Asp Pro Glu Asn Val Thr Ile Glu Ala
 435 440 445
 Pro Ser Ala Ser Arg Val Glu Leu Gly Ala Asp Arg Asn Ser His Ser
 450 455 460
 Ala Glu Val Ile Lys Val Thr Leu Lys Lys Asn Asn Thr Ser Leu Thr
 465 470 475 480
 Thr Leu Thr Asn Thr Thr Ile Ser Asn Leu Leu Lys Ser Ala His Val
 485 490 495
 Val Asn Ile Thr Ala Arg Arg Lys Leu Thr Val Asn Ser Ser Ile Ser
 500 505 510
 Ile Glu Arg Gly Ser His Leu Ile Leu His Ser Glu Gly Gln Gly Gly
 515 520 525
 Gln Gly Val Gln Ile Asp Lys Asp Ile Thr Ser Glu Gly Gly Asn Leu
 530 535 540
 Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn Ile Thr Leu
 545 550 555 560
 Gly Ser Gly Phe Leu Asn Ile Thr Thr Lys Glu Gly Asp Ile Ala Phe
 565 570 575
 Glu Asp Lys Ser Gly Arg Asn Asn Leu Thr Ile Thr Ala Gln Gly Thr
 580 585 590
 Ile Thr Ser Gly Asn Ser Asn Gly Phe Arg Phe Asn Asn Val Ser Leu
 595 600 605
 Asn Ser Leu Gly Gly Lys Leu Ser Phe Thr Asp Ser Arg Glu Asp Arg
 610 615 620
 Gly Arg Arg Thr Lys Gly Asn Ile Ser Asn Lys Phe Asp Gly Thr Leu
 625 630 635 640
 Asn Ile Ser Gly Thr Val Asp Ile Ser Met Lys Ala Pro Lys Val Ser
 645 650 655
 Trp Phe Tyr Arg Asp Lys Gly Arg Thr Tyr Trp Asn Val Thr Thr Leu
 660 665 670
 Asn Val Thr Ser Gly Ser Lys Phe Asn Leu Ser Ile Asp Ser Thr Gly
 675 680 685

95

Ser Gly Ser Thr Gly Pro Ser Ile Arg Asn Ala Glu Leu Asn Gly Ile
 690 695 700
 Thr Phe Asn Lys Ala Thr Phe Asn Ile Ala Gln Gly Ser Thr Ala Asn
 705 710 715 720
 Phe Ser Ile Lys Ala Ser Ile Met Pro Phe Lys Ser Asn Ala Asn Tyr
 725 730 735
 Ala Leu Phe Asn Glu Asp Ile Ser Val Ser Gly Gly Gly Ser Val Asn
 740 745 750
 Phe Lys Leu Asn Ala Ser Ser Ser Asn Ile Gln Thr Pro Gly Val Ile
 755 760 765
 Ile Lys Ser Gln Asn Phe Asn Val Ser Gly Gly Ser Thr Leu Asn Leu
 770 775 780
 Lys Ala Glu Gly Ser Thr Glu Thr Ala Phe Ser Ile Glu Asn Asp Leu
 785 790 795 800
 Asn Leu Asn Ala Thr Gly Gly Asn Ile Thr Ile Arg Gln Val Glu Gly
 805 810 815
 Thr Asp Ser Arg Val Asn Lys Gly Val Ala Ala Lys Lys Asn Ile Thr
 820 825 830
 Phe Lys Gly Gly Asn Ile Thr Phe Gly Ser Gln Lys Ala Thr Thr Glu
 835 840 845
 Ile Lys Gly Asn Val Thr Ile Asn Lys Asn Thr Asn Ala Thr Leu Arg
 850 855 860
 Gly Ala Asn Phe Ala Glu Asn Lys Ser Pro Leu Asn Ile Ala Gly Asn
 865 870 875 880
 Val Ile Asn Asn Gly Asn Leu Thr Thr Ala Gly Ser Ile Ile Asn Ile
 885 890 895
 Ala Gly Asn Leu Thr Val Ser Lys Gly Ala Asn Leu Gln Ala Ile Thr
 900 905 910
 Asn Tyr Thr Phe Asn Val Ala Gly Ser Phe Asp Asn Asn Gly Ala Ser
 915 920 925
 Asn Ile Ser Ile Ala Arg Gly Gly Ala Lys Phe Lys Asp Ile Asn Asn
 930 935 940
 Thr Ser Ser Leu Asn Ile Thr Thr Asn Ser Asp Thr Thr Tyr Arg Thr
 945 950 955 960
 Ile Ile Lys Gly Asn Ile Ser Asn Lys Ser Gly Asp Leu Asn Ile Ile
 965 970 975
 Asp Lys Lys Ser Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser Gln
 980 985 990
 Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Val Asn Ile Thr Asn
 995 1000 1005
 Gln Ile Thr Ile Lys Ala Gly Val Glu Gly Gly Arg Ser Asp Ser Ser
 1010 1015 1020
 Glu Ala Glu Asn Ala Asn Leu Thr Ile Gln Thr Lys Glu Leu Lys Leu
 1025 1030 1035 1040

Ala Gly Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr Ala
1045 1050 1055

Lys Asn Gly Ser Asp Leu Thr Ile Gly Asn Ala Ser Gly Gly Asn Ala
1060 1065 1070

Asp Ala Lys Lys Val Thr Phe Asp Lys Val Lys Asp Ser Lys Ile Ser
1075 1080 1085

Thr Asp Gly His Asn Val Thr Leu Asn Ser Glu Val Lys Thr Ser Asn
1090 1095 1100

Gly Ser Ser Asn Ala Gly Asn Asp Asn Ser Thr Gly Leu Thr Ile Ser
1105 1110 1115 1120

Ala Lys Asp Val Thr Val Asn Asn Asn Val Thr Ser His Lys Thr Ile
1125 1130 1135

Asn Ile Ser Ala Ala Ala Gly Asn Val Thr Thr Lys Glu Gly Thr Thr
1140 1145 1150

Ile Asn Ala Thr Thr Gly Ser Val Glu Val Thr Ala Gln Asn Gly Thr
1155 1160 1165

Ile Lys Gly Asn Ile Thr Ser Gln Asn Val Thr Val Thr Ala Thr Glu
1170 1175 1180

Asn Leu Val Thr Thr Glu Asn Ala Val Ile Asn Ala Thr Ser Gly Thr
1185 1190 1195 1200

Val Asn Ile Ser Thr Lys Thr Gly Asp Ile Lys Gly Gly Ile Glu Ser
1205 1210 1215

Thr Ser Gly Asn Val Asn Ile Thr Ala Ser Gly Asn Thr Leu Lys Val
1220 1225 1230

Ser Asn Ile Thr Gly Gln Asp Val Thr Val Thr Ala Asp Ala Gly Ala
1235 1240 1245

Leu Thr Thr Thr Ala Gly Ser Thr Ile Ser Ala Thr Thr Gly Asn Ala
1250 1255 1260

Asn Ile Thr Thr Lys Thr Gly Asp Ile Asn Gly Lys Val Glu Ser Ser
1265 1270 1275 1280

Ser Gly Ser Val Thr Leu Val Ala Thr Gly Ala Thr Leu Ala Val Gly
1285 1290 1295

Asn Ile Ser Gly Asn Thr Val Thr Ile Thr Ala Asp Ser Gly Lys Leu
1300 1305 1310

Thr Ser Thr Val Gly Ser Thr Ile Asn Gly Thr Asn Ser Val Thr Thr
1315 1320 1325

Ser Ser Gln Ser Gly Asp Ile Glu Gly Thr Ile Ser Gly Asn Thr Val
1330 1335 1340

Asn Val Thr Ala Ser Thr Gly Asp Leu Thr Ile Gly Asn Ser Ala Lys
1345 1350 1355 1360

Val Glu Ala Lys Asn Gly Ala Ala Thr Leu Thr Ala Glu Ser Gly Lys
1365 1370 1375

Leu Thr Thr Gln Thr Gly Ser Ser Ile Thr Ser Ser Asn Gly Gln Thr
1380 1385 1390

97

Thr Leu Thr Ala Lys Asp Ser Ser Ile Ala Gly Asn Ile Asn Ala Ala
 1395 1400 1405
 Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Thr Gly Asp Ser
 1410 1415 1420
 Lys Ile Asn Ala Thr Ser Gly Thr Leu Thr Ile Asn Ala Lys Asp Ala
 1425 1430 1435 1440
 Lys Leu Asp Gly Ala Ala Ser Gly Asp Arg Thr Val Val Asn Ala Thr
 1445 1450 1455
 Asn Ala Ser Gly Ser Gly Asn Val Thr Ala Lys Thr Ser Ser Ser Val
 1460 1465 1470
 Asn Ile Thr Gly Asp Leu Asn Thr Ile Asn Gly Leu Asn Ile Ile Ser
 1475 1480 1485
 Glu Asn Gly Arg Asn Thr Val Arg Leu Arg Gly Lys Glu Ile Asp Val
 1490 1495 1500
 Lys Tyr Ile Gln Pro Gly Val Ala Ser Val Glu Glu Val Ile Glu Ala
 1505 1510 1515 1520
 Lys Arg Val Leu Glu Lys Val Lys Asp Leu Ser Asp Glu Glu Arg Glu
 1525 1530 1535
 Thr Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Val Glu Pro Asn
 1540 1545 1550
 Asn Ala Ile Thr Val Asn Thr Gln Asn Glu Phe Thr Thr Lys Pro Ser
 1555 1560 1565
 Ser Gln Val Thr Ile Ser Glu Gly Lys Ala Cys Phe Ser Ser Gly Asn
 1570 1575 1580
 Gly Ala Arg Val Cys Thr Asn Val Ala Asp Asp Gly Gln Gln Pro
 1585 1590 1595

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1600 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu
 1 5 10 15
 Val Ala Val Ser Glu Leu Thr Arg Gly Cys Asp His Ser Thr Glu Lys
 20 25 30
 Gly Ser Glu Lys Pro Val Arg Thr Lys Val Arg His Leu Ala Leu Lys
 35 40 45
 Pro Leu Ser Ala Ile Leu Leu Ser Leu Gly Met Ala Ser Ile Pro Gln
 50 55 60
 Ser Val Leu Ala Ser Gly Leu Gln Gly Met Ser Val Val His Gly Thr
 65 70 75 80

Ala Thr Met Gln Val Asp Gly Asn Lys Thr Thr Ile Arg Asn Ser Val
85 90 95
Asn Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met
100 105 110
Glu Gln Phe Leu Gln Glu Ser Ser Asn Ser Ala Val Phe Asn Arg Val
115 120 125
Thr Ser Asp Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly
130 135 140
Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
145 150 155 160
Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
165 170 175
Glu Asn Ile Lys Ala Arg Asn Phe Thr Leu Glu Gln Thr Lys Asp Lys
180 185 190
Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
195 200 205
Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
210 215 220
Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
225 230 235 240
Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
245 250 255
Glu Asn Glu Ala Ile Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn
260 265 270
Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Lys Gly Lys Leu Ser Ala
275 280 285
Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
290 295 300
Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
305 310 315 320
Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
325 330 335
Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
340 345 350
Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala
355 360 365
Lys Lys Thr Thr Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
370 375 380
Glu Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
385 390 395 400
Gly Asn Ile Asn Ala Gln Gly Ser Asp Ile Ala Lys Thr Gly Gly Phe
405 410 415
Val Glu Thr Ser Gly His Asp Leu Ser Ile Gly Asp Asp Val Ile Val
420 425 430

Asp Ala Lys Glu Trp Leu Leu Asp Pro Asp Asp Val Ser Ile Glu Thr
 435 440 445
 Leu Thr Ser Gly Arg Asn Asn Thr Gly Glu Asn Gln Gly Tyr Thr Thr
 450 455 460
 Gly Asp Gly Thr Lys Glu Ser Pro Lys Gly Asn Ser Ile Ser Lys Pro
 465 470 475 480
 Thr Leu Thr Asn Ser Thr Leu Glu Gln Ile Leu Arg Arg Gly Ser Tyr
 485 490 495
 Val Asn Ile Thr Ala Asn Asn Arg Ile Tyr Val Asn Ser Ser Ile Asn
 500 505 510
 Leu Ser Asn Gly Ser Leu Thr Leu His Thr Lys Arg Asp Gly Val Lys
 515 520 525
 Ile Asn Gly Asp Ile Thr Ser Asn Glu Asn Gly Asn Leu Thr Ile Lys
 530 535 540
 Ala Gly Ser Trp Val Asp Val His Lys Asn Ile Thr Leu Gly Thr Gly
 545 550 555 560
 Phe Leu Asn Ile Val Ala Gly Asp Ser Val Ala Phe Glu Arg Glu Gly
 565 570 575
 Asp Lys Ala Arg Asn Ala Thr Asp Ala Gln Ile Thr Ala Gln Gly Thr
 580 585 590
 Ile Thr Val Asn Lys Asp Asp Lys Gln Phe Arg Phe Asn Asn Val Ser
 595 600 605
 Leu Asn Gly Thr Gly Lys Gly Leu Lys Phe Ile Ala Asn Gln Asn Asn
 610 615 620
 Phe Thr His Lys Phe Asp Gly Glu Ile Asn Ile Ser Gly Ile Val Thr
 625 630 635 640
 Ile Asn Gln Thr Thr Lys Lys Asp Val Lys Tyr Trp Asn Ala Ser Lys.
 645 650 655
 Asp Ser Tyr Trp Asn Val Ser Ser Leu Thr Leu Asn Thr Val Gln Lys
 660 665 670
 Phe Thr Phe Ile Lys Phe Val Asp Ser Gly Ser Asn Gly Gln Asp Leu
 675 680 685
 Arg Ser Ser Arg Arg Ser Phe Ala Gly Val His Phe Asn Gly Ile Gly
 690 695 700
 Gly Lys Thr Asn Phe Asn Ile Gly Ala Asn Ala Lys Ala Leu Phe Lys
 705 710 715 720
 Leu Lys Pro Asn Ala Ala Thr Asp Pro Lys Lys Glu Leu Pro Ile Thr
 725 730 735
 Phe Asn Ala Asn Ile Thr Ala Thr Gly Asn Ser Asp Ser Ser Val Met
 740 745 750
 Phe Asp Ile His Ala Asn Leu Thr Ser Arg Ala Ala Gly Ile Asn Met
 755 760 765
 Asp Ser Ile Asn Ile Thr Gly Gly Leu Asp Phe Ser Ile Thr Ser His
 770 775 780

100

Asn Arg Asn Ser Asn Ala Phe Glu Ile Lys Lys Asp Leu Thr Ile Asn
 785 790 795 800
 Ala Thr Gly Ser Asn Phe Ser Leu Lys Gln Thr Lys Asp Ser Phe Tyr
 805 810 815
 Asn Glu Tyr Ser Lys His Ala Ile Asn Ser Ser His Asn Leu Thr Ile
 820 825 830
 Leu Gly Gly Asn Val Thr Leu Gly Gly Glu Asn Ser Ser Ser Ile
 835 840 845
 Thr Gly Asn Ile Asn Ile Thr Asn Lys Ala Asn Val Thr Leu Gln Ala
 850 855 860
 Asp Thr Ser Asn Ser Asn Thr Gly Leu Lys Lys Arg Thr Leu Thr Leu
 865 870 875 880
 Gly Asn Ile Ser Val Glu Gly Asn Leu Ser Leu Thr Gly Ala Asn Ala
 885 890 895
 Asn Ile Val Gly Asn Leu Ser Ile Ala Glu Asp Ser Thr Phe Lys Gly
 900 905 910
 Glu Ala Ser Asp Asn Leu Asn Ile Thr Gly Thr Phe Thr Asn Asn Gly
 915 920 925
 Thr Ala Asn Ile Asn Ile Lys Gly Val Val Lys Leu Gly Asp Ile Asn
 930 935 940
 Asn Lys Gly Gly Leu Asn Ile Thr Thr Asn Ala Ser Gly Thr Gln Lys
 945 950 955 960
 Thr Ile Ile Asn Gly Asn Ile Thr Asn Glu Lys Gly Asp Leu Asn Ile
 965 970 975
 Lys Asn Ile Lys Ala Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser
 980 985 990
 Gln Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Val Asn Ile Thr
 995 1000 1005
 Asn Gln Ile Thr Ile Lys Ala Gly Val Glu Gly Gly Arg Ser Asp Ser
 1010 1015 1020
 Ser Glu Ala Glu Asn Ala Asn Leu Thr Ile Gln Thr Lys Glu Leu Lys
 1025 1030 1035 1040
 Leu Ala Gly Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr
 1045 1050 1055
 Ala Lys Asn Gly Ser Asp Leu Thr Ile Gly Asn Ala Ser Gly Gly Asn
 1060 1065 1070
 Ala Asp Ala Lys Lys Val Thr Phe Asp Lys Val Lys Asp Ser Lys Ile
 1075 1080 1085
 Ser Thr Asp Gly His Asn Val Thr Leu Asn Ser Glu Val Lys Thr Ser
 1090 1095 1100
 Asn Gly Ser Ser Asn Ala Gly Asn Asp Asn Ser Thr Gly Leu Thr Ile
 1105 1110 1115 1120
 Ser Ala Lys Asp Val Thr Val Asn Asn Asn Val Thr Ser His Lys Thr
 1125 1130 1135

101

Ile Asn Ile Ser Ala Ala Ala Gly Asn Val Thr Thr Lys Glu Gly Thr
1140 1145 1150

Thr Ile Asn Ala Thr Thr Gly Ser Val Glu Val Thr Ala Gln Asn Gly
1155 1160 1165

Thr Ile Lys Gly Asn Ile Thr Ser Gln Asn Val Thr Val Thr Ala Thr
1170 1175 1180

Glu Asn Leu Val Thr Thr Glu Asn Ala Val Ile Asn Ala Thr Ser Gly
1185 1190 1195 1200

Thr Val Asn Ile Ser Thr Lys Thr Gly Asp Ile Lys Gly Gly Ile Glu
1205 1210 1215

Ser Thr Ser Gly Asn Val Asn Ile Thr Ala Ser Gly Asn Thr Leu Lys
1220 1225 1230

Val Ser Asn Ile Thr Gly Gln Asp Val Thr Val Thr Ala Asp Ala Gly
1235 1240 1245

Ala Leu Thr Thr Thr Ala Gly Ser Thr Ile Ser Ala Thr Thr Gly Asn
1250 1255 1260

Ala Asn Ile Thr Thr Lys Thr Gly Asp Ile Asn Gly Lys Val Glu Ser
1265 1270 1275 1280

Ser Ser Gly Ser Val Thr Leu Val Ala Thr Gly Ala Thr Leu Ala Val
1285 1290 1295

Gly Asn Ile Ser Gly Asn Thr Val Thr Ile Thr Ala Asp Ser Gly Lys
1300 1305 1310

Leu Thr Ser Thr Val Gly Ser Thr Ile Asn Gly Thr Asn Ser Val Thr
1315 1320 1325

Thr Ser Ser Gln Ser Gly Asp Ile Glu Gly Thr Ile Ser Gly Asn Thr
1330 1335 1340

Val Asn Val Thr Ala Ser Thr Gly Asp Leu Thr Ile Gly Asn Ser Ala
1345 1350 1355 1360

Lys Val Glu Ala Lys Asn Gly Ala Ala Thr Leu Thr Ala Glu Ser Gly
1365 1370 1375

Lys Leu Thr Thr Gln Thr Gly Ser Ser Ile Thr Ser Ser Asn Gly Gln
1380 1385 1390

Thr Thr Leu Thr Ala Lys Asp Ser Ser Ile Ala Gly Asn Ile Asn Ala
1395 1400 1405

Ala Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Thr Gly Asp
1410 1415 1420

Ser Lys Ile Asn Ala Thr Ser Gly Thr Leu Thr Ile Asn Ala Lys Asp
1425 1430 1435 1440

Ala Lys Leu Asp Gly Ala Ala Ser Gly Asp Arg Thr Val Val Asn Ala
1445 1450 1455

Thr Asn Ala Ser Gly Ser Gly Asn Val Thr Ala Lys Thr Ser Ser Ser
1460 1465 1470

Val Asn Ile Thr Gly Asp Leu Asn Thr Ile Asn Gly Leu Asn Ile Ile
1475 1480 1485

102

Ser Glu Asn Gly Arg Asn Thr Val Arg Leu Arg Gly Lys Glu Ile Asp
 1490 1495 1500
 Val Lys Tyr Ile Gln Pro Gly Val Ala Ser Val Glu Glu Val Ile Glu
 1505 1510 1515 1520
 Ala Lys Arg Val Leu Glu Lys Val Lys Asp Leu Ser Asp Glu Glu Arg
 1525 1530 1535
 Glu Thr Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Val Glu Pro
 1540 1545 1550
 Asn Asn Ala Ile Thr Val Asn Thr Gln Asn Glu Phe Thr Thr Lys Pro
 1555 1560 1565
 Ser Ser Gln Val Thr Ile Ser Glu Gly Lys Ala Cys Phe Ser Ser Gly
 1570 1575 1580
 Asn Gly Ala Arg Val Cys Thr Asn Val Ala Asp Asp Gly Gln Gln Pro
 1585 1590 1595 1600

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 29 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Val Asp Glu Val Ile Glu Ala Lys Arg Ile Leu Glu Lys Val Lys Asp
 1 5 10 15
 Leu Ser Asp Glu Glu Arg Glu Ala Leu Ala Lys Leu Gly
 20 25

SEQUENCE LISTING

<110> Barenkamp, Stephen J.

<120> HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
HAEMOPHILUS

<130>

<140>

<141>

<150> 09/155,614

<151> 1998-09-30

<150> 08/617,697

<151> 1996-04-01

<150> PCT/US97/04707

<151> 1997-04-01

<160> 11

<170> PatentIn Ver. 2.1

<210> 1

<211> 5116

<212> DNA

<213> Haemophilus influenzae

<400> 1

```

acagcgcttct cttaatacta gtacaaaccc acaataaaa atgacaaaa acaattacaa 60
cacctttttt gcagtcctata tgcaaatatt ttaaaaaata gtataaatcc gccatataaa 120
atggtataaat ctctcatctt tcctctttca tccttcacat ttcatctttc atcttttcac 180
ttctcatctt catctttcat ctcttcacat tcctctttca tccttcacat ttcatctttc 240
acatgccttg atgaaccgag ggaagggagg gaggggcaag aatgaagagg gagctgaacg 300
aacgcaaatg ataaagtaat ttaattgttc aactaacctt aggagaaaa atgaacaagc 360
tatatcgctc caaatccagc aaacgcctga atgctttggg tgctgtgctt gaattggcac 420
ggggttgtag ccattccaca gaaaaaggca gcgaaaaacc tgctcgcatg aaagtgcgtc 480
acttagcggt aaagccactt tccgctatgt tactatcttt aggtgtaaca tctattccac 540
aatctgtttt agcaagcggc ttacaaggaa tggatgtagt acacggcaca gccactatgc 600
aagttagatg taataaaacc attatccgca acagtgttga cgatatcatt aattggaacc 660
aatttaacat cgaccaaaaa gaaatgggtc agtttttaca agaaaaaac aactccgcgc 720
tattcaacgc tgtttacatc aaccaaactc cccaattaaa agggatttta gattctaacg 780
gacacgctct ttttaacca ccaaatggta tcacaatagg taaagacgca attattaaca 840
ctaattggct tacgcttctc acgctagaca tttctaacga aaacatcaag gcgcgtaatt 900
tcaccttcga gcaaacccaa gataaagcgc tcgctgaaat tgtgaatcac ggtttaatta 960
ctgtcggtta agacggcagt gtaaatctta ttggtggcaa agtgaaaaac gaggggtgta 1020
ttagcgtaaa tgggtggcagc atttctttac tcgcagggca aaaaatcacc atcagcgata 1080
taataaaccc aaccattact tacagcattg ccgcgcctga aatggaagcg gtcaaatctgg 1140
cgcatatttt tgcacaaaggc ggtaacatta atgtccgtgc tgccactatt cgaaaccacg 1200
gtaaactttt tctgtattct gtaagcaaa caataatgtt ctcttcgcga ctttcgcgca 1260
aagaggggtg agcggaattt ggcggtgtaa tttccgtca aaatcagcaa gctaaaggcg 1320
gcagctgatg gattcacagg gataaagtca cattaaaaac aggtgcagtt atcgaccttt 1380
caggtaaaag agggggagaa acttaacctg gcggtgacga ccgcggcgaa ggtaaaaagg 1440
gcattcaatt agcaaaagaa acctctttag aaaaaggctc aacctcaat gtatcaggca 1500
aagaaaaaag cggacgcgct attgtgtggg gcgatattgc gtaattagc ggcaatatta 1560
acgctcaagg tagtggtgat atcgctaaaa ccggtgtgtt tgtggagacg tcggggcgatg 1620
atttattcat caaagcaaat gcaattgttg accgcaaaaga gtggttgta gacccgata 1680
atgtatctat taatgcagaa acagcaggac gcagcaatac ttcagaagac gatgaatata 1740
cgggataccg gaatagtgc agcaccceaa aacgaaacaa agaaaagaca acattacaaa 1800

```

acacaactct	tgagagtata	ctaaaaaaag	gtacctttgt	taaacatcact	gctaatacaac	1860
gcatactatgt	caatagctcc	attaatttat	ccaatggcag	cttaactctct	tgagtgagg	1920
gtcggagcgg	tgccggcggt	gagattaaca	acgatattac	caccggtgat	gataccagag	1980
gtgcaaacct	aacaatttac	tcaggcggt	gggttgatgt	tcataaaaa	atctcaactcg	2040
ggggcgtaag	tacacataaac	attacagcta	aacaagatat	cgctttttag	aaaggaagca	2100
accaagtcat	tacaggtcaa	gggactatta	cctcaggcaa	tcaaaaaggt	tttagattta	2160
ataatgtctc	cttaaacggc	actggcagcg	gactgcaatt	caccactaaa	agaaccaata	2220
atacgcgtat	cacaaaaata	tttgaaggga	cttaaaat	ttcagggaaa	gtgaactatc	2280
caatggtttt	acctaaaaat	gaagtggtat	atgataaatt	caaaaggcgc	acttaactga	2340
atttaacctc	cttaaaattgt	tccgagagtg	cgagttttaa	cctcactatt	gactccagag	2400
gaagcgatag	tgacggcaca	cttaccgcgc	cttataaatt	aaacggtata	ctattcaaca	2460
aagacactac	ctttaattgt	gaacgaaatg	caagagtcaa	ctttgacatc	aaggaccaa	2520
tagggataaa	taagtattct	agtttgaatt	acgcatcatt	taatggaaac	atttcagttt	2580
cgggaggggg	gagtggtgat	ttcacacttc	tcgctctate	ctctaactgc	caaacccccc	2640
gtgtagttat	aaattctaaa	tactttaatg	tttcaacagg	gtcaagttaa	agatttaaaa	2700
cttcagcctc	aacaaaaact	ggcttctcaa	tagagaaa	tttaacttta	aatgccacgc	2760
gaggcaacat	aacacttttg	caagttgaag	gcaccgatgg	aatgattgtg	aaaggcattg	2820
tagccaaaaa	aaacataacc	tttgaaggag	gtaacatcac	ctttggctcc	aggaaagcgc	2880
taacagaaat	cgaaggcaat	gttactatca	atacaaacgc	taacgtcac	cttatcggtt	2940
cggattttga	cagacatcaa	aaacctttaa	ctattaaaaa	agatgtcatc	atttaagtgc	3000
gcacacctac	cgctggagct	aatattgtca	atatagccgg	aaatcttacc	gttgaagta	3060
acgctaattt	caaagctatc	acaaatttca	cttttaatgt	aggcggtctg	tttgacaaac	3120
aaggcaattc	aaatatttcc	attgcaaaag	gaggggctcg	ctttaaagac	attgataatt	3180
ccaagaattt	aaagctaccc	accaactcca	gtccactatt	cgcacattat	atcagggcca	3240
atatacccaa	taaaaaagg	gatttaataa	ttacgaacga	aggtagtgat	actgaatagc	3300
aaattggcgg	cgatgtctcg	caaaaagaag	gtaatctcac	gatttctctc	gacaaaatca	3360
atattaccaa	acagataaca	atcaaggcag	gtgttgatgg	ggagaattcc	gattcagagc	3420
cgacaaaaca	tggcaatcta	accattaaaa	caaagaattg	gaaattaaag	caagacctaa	3480
atatttcagg	tttcaataaa	gcagagatta	cagctaaaaga	tggtagtgat	ttactattgt	3540
gtaacaccaa	tagtgctgat	ggtaactaat	ccaaaaaagt	aacctttaac	cagggttaag	3600
attcaaaaat	ctctgctgac	ggtcacaaag	tgacactaca	cagcaaaagt	gaacatccg	3660
gtagtataaa	caacactgaa	gatagcagtg	acaataatgc	cggtcttaact	atcgatgcaa	3720
aaaatgtaaa	agtaaaaac	aatattactt	ctcacaaag	agtgaagcatc	tctgcgacaa	3780
gtgggaaat	taccactaaa	acaggtaaca	ccattaaacg	aaccactggt	aacgtggaga	3840
taaccgctca	aacaggtagt	atcctagggt	gaattgagtc	cagctctggc	tctgtaacac	3900
ttactgcaac	cgagggcgct	cttgcgttaa	gcaatttttc	gggcaacacc	gttactgtta	3960
ctgcaaatag	cggctgcatta	accactttgg	caggctctac	aatttaaggga	accgaagtg	4020
taaccacttc	aagtcaatca	ggcgataatc	gcggtacgat	ttctggtggc	acagttaggg	4080
ttaaagcaac	cgaagttta	accactcaat	ccaattcaaa	aattaaagca	acaacaggcg	4140
aggctcaactg	aacaagtgca	acagggtacaa	ttggtggtag	gatttccggg	aatacggtaa	4200
atgttaccgc	aaacgctggc	gatttaacag	ttgggaatgg	cgagaaaatt	aatgcgacag	4260
aaggagctgc	acccttaact	acatcatcgg	gcaaattaac	taccgaagct	agttcacaca	4320
ttacttcagc	caagggtcag	gtaaatcttt	cagctcagga	tggtagcggt	gcagggaagta	4380
taattgcgc	caatgtgaca	ctaaatacta	caggcacttt	actacocgtc	aaaggttcaa	4440
acattaatgc	aaccagcggt	accttgggtta	ttacgcgaaa	agacgctgag	ctaaattggc	4500
cagacttggg	taaccacaca	gtggttaattg	aaacgaacgc	aaatggctcc	ggcagcgtaa	4560
tcgcgacac	ctcaagcaga	gtgaacatca	ctgggggattt	aatcacataa	aatggattaa	4620
atatcatctt	aaaaaacggg	ataaacacgc	tactgtttaa	aggcggttaa	attgatgtga	4680
atacatctta	accgggtata	gcaagcgtag	atgaagtaat	tgaagcgaaa	cgcactcttg	4740
agaaggtaaa	agatttatct	gatgaagaaa	gagaagcgat	agctaaacct	ggagtttaagt	4800
gttgactgtt	tattgacaca	aataatacaa	ttacagctga	tacacaaaat	gaattgtcaa	4860
cgagaccatt	aagtcgaata	gtgattttct	aaggcagggc	gtgtttctca	aacagtgaatg	4920
gcgcagcgg	gtcgtttaat	atcgcgtgata	acggcggtga	cgcgctagta	attgacaagg	4980
tagattttcat	cctgcaatga	agtcattttta	ttttcgtatt	atttactgtg	tggtttaaag	5040
ttcagtcagg	gctttaccaca	tcttgtaaaa	aattacggag	aatacaataa	agtattttta	5100
acaggttatt	attatg					5116

<210> 2
 <211> 1536
 <212> PRT
 <213> Haemophilus influenzae

<400> 2

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu
 1 5 10 15

Val Ala Val Ser Glu Leu Ala Arg Gly Cys Asp His Ser Thr Glu Lys
 20 25 30

Gly Ser Glu Lys Pro Ala Arg Met Lys Val Arg His Leu Ala Leu Lys
 35 40 45

Pro Leu Ser Ala Met Leu Leu Ser Leu Gly Val Thr Ser Ile Pro Gln
 50 55 60

Ser Val Leu Ala Ser Gly Leu Gln Gly Met Asp Val Val His Gly Thr
 65 70 75 80

Ala Thr Met Gln Val Asp Gly Asn Lys Thr Ile Ile Arg Asn Ser Val
 85 90 95

Asp Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met
 100 105 110

Val Gln Phe Leu Gln Glu Asn Asn Asn Ser Ala Val Phe Asn Arg Val
 115 120 125

Thr Ser Asn Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly
 130 135 140

Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
 145 150 155 160

Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
 165 170 175

Glu Asn Ile Lys Ala Arg Asn Phe Thr Phe Glu Gln Thr Lys Asp Lys
 180 185 190

Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
 195 200 205

Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
 210 215 220

Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
 225 230 235 240

Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
 245 250 255

Glu Asn Glu Ala Val Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn
 260 265 270

Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Gln Gly Lys Leu Ser Ala
 275 280 285

Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
 290 295 300
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
 305 310 315 320
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
 325 330 335
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
 340 345 350
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala
 355 360 365
 Lys Lys Thr Ser Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
 370 375 380
 Glu Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
 385 390 395 400
 Gly Asn Ile Asn Ala Gln Gly Ser Gly Asp Ile Ala Lys Thr Gly Gly
 405 410 415
 Phe Val Glu Thr Ser Gly His Asp Leu Phe Ile Lys Asp Asn Ala Ile
 420 425 430
 Val Asp Ala Lys Glu Trp Leu Leu Asp Phe Asp Asn Val Ser Ile Asn
 435 440 445
 Ala Glu Thr Ala Gly Arg Ser Asn Thr Ser Glu Asp Asp Glu Tyr Thr
 450 455 460
 Gly Ser Gly Asn Ser Ala Ser Thr Pro Lys Arg Asn Lys Glu Lys Thr
 465 470 475 480
 Thr Leu Thr Asn Thr Thr Leu Glu Ser Ile Leu Lys Lys Gly Thr Phe
 485 490 495
 Val Asn Ile Thr Ala Asn Gln Arg Ile Tyr Val Asn Ser Ser Ile Asn
 500 505 510
 Leu Ser Asn Gly Ser Leu Thr Leu Trp Ser Glu Gly Arg Ser Gly Gly
 515 520 525
 Gly Val Glu Ile Asn Asn Asp Ile Thr Thr Gly Asp Asp Thr Arg Gly
 530 535 540
 Ala Asn Leu Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn
 545 550 555 560
 Ile Ser Leu Gly Ala Gln Gly Asn Ile Asn Ile Thr Ala Lys Gln Asp
 565 570 575
 Ile Ala Phe Glu Lys Gly Ser Asn Gln Val Ile Thr Gly Gln Gly Thr
 580 585 590

Ile Thr Ser Gly Asn Gln Lys Gly Phe Arg Phe Asn Asn Val Ser Leu
 595 600 605
 Asn Gly Thr Gly Ser Gly Leu Gln Phe Thr Thr Lys Arg Thr Asn Lys
 610 615 620
 Tyr Ala Ile Thr Asn Lys Phe Glu Gly Thr Leu Asn Ile Ser Gly Lys
 625 630 635 640
 Val Asn Ile Ser Met Val Leu Pro Lys Asn Glu Ser Gly Tyr Asp Lys
 645 650 655
 Phe Lys Gly Arg Thr Tyr Trp Asn Leu Thr Ser Leu Asn Val Ser Glu
 660 665 670
 Ser Gly Glu Phe Asn Leu Thr Ile Asp Ser Arg Gly Ser Asp Ser Ala
 675 680 685
 Gly Thr Leu Thr Gln Pro Tyr Asn Leu Asn Gly Ile Ser Phe Asn Lys
 690 695 700
 Asp Thr Thr Phe Asn Val Glu Arg Asn Ala Arg Val Asn Phe Asp Ile
 705 710 715 720
 Lys Ala Pro Ile Gly Ile Asn Lys Tyr Ser Ser Leu Asn Tyr Ala Ser
 725 730 735
 Phe Asn Gly Asn Ile Ser Val Ser Gly Gly Gly Ser Val Asp Phe Thr
 740 745 750
 Leu Leu Ala Ser Ser Ser Asn Val Gln Thr Pro Gly Val Val Ile Asn
 755 760 765
 Ser Lys Tyr Phe Asn Val Ser Thr Gly Ser Ser Leu Arg Phe Lys Thr
 770 775 780
 Ser Gly Ser Thr Lys Thr Gly Phe Ser Ile Glu Lys Asp Leu Thr Leu
 785 790 795 800
 Asn Ala Thr Gly Gly Asn Ile Thr Leu Leu Gln Val Glu Gly Thr Asp
 805 810 815
 Gly Met Ile Gly Lys Gly Ile Val Ala Lys Lys Asn Ile Thr Phe Glu
 820 825 830
 Gly Gly Asn Ile Thr Phe Gly Ser Arg Lys Ala Val Thr Glu Ile Glu
 835 840 845
 Gly Asn Val Thr Ile Asn Asn Asn Ala Asn Val Thr Leu Ile Gly Ser
 850 855 860
 Asp Phe Asp Asn His Gln Lys Pro Leu Thr Ile Lys Lys Asp Val Ile
 865 870 875 880
 Ile Asn Ser Gly Asn Leu Thr Ala Gly Gly Asn Ile Val Asn Ile Ala
 885 890 895
 Gly Asn Leu Thr Val Glu Ser Asn Ala Asn Phe Lys Ala Ile Thr Asn
 900 905 910

Phe Thr Phe Asn Val Gly Gly Leu Phe Asp Asn Lys Gly Asn Ser Asn
 915 920 925
 Ile Ser Ile Ala Lys Gly Gly Ala Arg Phe Lys Asp Ile Asp Asn Ser
 930 935 940
 Lys Asn Leu Ser Ile Thr Thr Asn Ser Ser Ser Thr Tyr Arg Thr Ile
 945 950 955 960
 Ile Ser Gly Asn Ile Thr Asn Lys Asn Gly Asp Leu Asn Ile Thr Asn
 965 970 975
 Glu Gly Ser Asp Thr Glu Met Gln Ile Gly Gly Asp Val Ser Gln Lys
 980 985 990
 Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Ile Asn Ile Thr Lys Gln
 995 1000 1005
 Ile Thr Ile Lys Ala Gly Val Asp Gly Glu Asn Ser Asp Ser Asp Ala
 1010 1015 1020
 Thr Asn Asn Ala Asn Leu Thr Ile Lys Thr Lys Glu Leu Lys Leu Thr
 1025 1030 1035 1040
 Gln Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr Ala Lys
 1045 1050 1055
 Asp Gly Ser Asp Leu Thr Ile Gly Asn Thr Asn Ser Ala Asp Gly Thr
 1060 1065 1070
 Asn Ala Lys Lys Val Thr Phe Asn Gln Val Lys Asp Ser Lys Ile Ser
 1075 1080 1085
 Ala Asp Gly His Lys Val Thr Leu His Ser Lys Val Glu Thr Ser Gly
 1090 1095 1100
 Ser Asn Asn Asn Thr Glu Asp Ser Ser Asp Asn Asn Ala Gly Leu Thr
 1105 1110 1115 1120
 Ile Asp Ala Lys Asn Val Thr Val Asn Asn Asn Ile Thr Ser His Lys
 1125 1130 1135
 Ala Val Ser Ile Ser Ala Thr Ser Gly Glu Ile Thr Thr Lys Thr Gly
 1140 1145 1150
 Thr Thr Ile Asn Ala Thr Thr Gly Asn Val Glu Ile Thr Ala Gln Thr
 1155 1160 1165
 Gly Ser Ile Leu Gly Gly Ile Glu Ser Ser Ser Gly Ser Val Thr Leu
 1170 1175 1180
 Thr Ala Thr Glu Gly Ala Leu Ala Val Ser Asn Ile Ser Gly Asn Thr
 1185 1190 1195 1200
 Val Thr Val Thr Ala Asn Ser Gly Ala Leu Thr Thr Leu Ala Gly Ser
 1205 1210 1215

Thr Ile Lys Gly Thr Glu Ser Val Thr Thr Ser Ser Gln Ser Gly Asp
 1220 1225 1230
 Ile Gly Gly Thr Ile Ser Gly Gly Thr Val Glu Val Lys Ala Thr Glu
 1235 1240 1245
 Ser Leu Thr Thr Gln Ser Asn Ser Lys Ile Lys Ala Thr Thr Gly Glu
 1250 1255 1260
 Ala Asn Val Thr Ser Ala Thr Gly Thr Ile Gly Gly Thr Ile Ser Gly
 1265 1270 1275 1280
 Asn Thr Val Asn Val Thr Ala Asn Ala Gly Asp Leu Thr Val Gly Asn
 1285 1290 1295
 Gly Ala Glu Ile Asn Ala Thr Glu Gly Ala Ala Thr Leu Thr Thr Ser
 1300 1305 1310
 Ser Gly Lys Leu Thr Thr Glu Ala Ser Ser His Ile Thr Ser Ala Lys
 1315 1320 1325
 Gly Gln Val Asn Leu Ser Ala Gln Asp Gly Ser Val Ala Gly Ser Ile
 1330 1335 1340
 Asn Ala Ala Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Val
 1345 1350 1355 1360
 Lys Gly Ser Asn Ile Asn Ala Thr Ser Gly Thr Leu Val Ile Asn Ala
 1365 1370 1375
 Lys Asp Ala Glu Leu Asn Gly Ala Ala Leu Gly Asn His Thr Val Val
 1380 1385 1390
 Asn Ala Thr Asn Ala Asn Gly Ser Gly Ser Val Ile Ala Thr Thr Ser
 1395 1400 1405
 Ser Arg Val Asn Ile Thr Gly Asp Leu Ile Thr Ile Asn Gly Leu Asn
 1410 1415 1420
 Ile Ile Ser Lys Asn Gly Ile Asn Thr Val Leu Leu Lys Gly Val Lys
 1425 1430 1435 1440
 Ile Asp Val Lys Tyr Ile Gln Pro Gly Ile Ala Ser Val Asp Gly Val
 1445 1450 1455
 Ile Glu Ala Lys Arg Ile Leu Glu Lys Val Lys Asp Leu Ser Asp Glu
 1460 1465 1470
 Glu Arg Glu Ala Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Ile
 1475 1480 1485
 Glu Pro Asn Asn Thr Ile Thr Val Asp Thr Gln Asn Glu Phe Ala Thr
 1490 1495 1500
 Arg Pro Leu Ser Arg Ile Val Ile Ser Glu Gly Arg Ala Cys Phe Ser
 1505 1510 1515 1520
 Asn Ser Asp Gly Ala Thr Val Cys Val Asn Ile Ala Asp Asn Gly Arg
 1525 1530 1535

<210> 3
 <211> 4937
 <212> DNA
 <213> Haemophilus influenzae

<400> 3
 taaatataca agataataaa aataaatcaa gatttttgtg atgacaaaaa acaattacaa 60
 cacctttttt gcagctctata tgcaaatatt ttaaaaaaat agtataaatc cgccatataa 120
 aatgggtataa tcttttcatt ttcattcttta atctttcatt tttcatcttt 180
 ctttcattct tcatctttca tcttttcatt tttcatcttt atctttcatt 240
 cacatgaaat gatgaaccga ggggaaggag ggagggggcaa gaatgaagag ggagctgaac 300
 gaacgcaa atgataaagtaa ttttaattgt caactaacct taggagaaaa tatgaacaag 360
 atatatcgtc tcaaatccag caaacgcctg aatgccttgg ttgcgtgtgc tgaattggca 420
 cggggttggt accattccac agaaaaaggc ttccgctatg ttactatctt taggtgtaac 480
 cacttagcgt taaagccact ttccgctatg ttactatctt taggtgtaac atctattcca 540
 caactctgtt tagcaaggcg cttacaagga atggatgtag tacacggcac agccactatg 600
 caagtagatg gtaataaaaa cattatccgc aacagtggtg acgctatcat taattggaaa 660
 caatttaaca tcgacacaaa tgaatgggtg cagtttttac aagaaaaaaa caactccgcc 720
 gtattcaacc gtgttacatc taaccaaatc tcccaattaa aagggtattt agattctaac 780
 ggacaagtct ttttaatcaa cccaaatggt atcaacaatg gtaaaagcgc aattattaac 840
 actaaaggct ttacggtctc tacgctagac atttctaagc aaaacatcaa ggcgcgtaat 900
 ttcccttctg agcaaaccaa agataaaggc ctgcgtgaaa ttgtgaatca cgggttaatt 960
 actgctcgta aagaacggcg tgaataatct attgtggcca aagtgaaaaa cgagggtgtg 1020
 attagcgtaa atggtggcag catttcttta ctgcagggcg aaaaaatcac catcagcgat 1080
 ataataaacc caaccattac ttacagcatt gccgcgcctg aaaaatgaagc ggtcaactct 1140
 ggcgataatt ttgcacaaagg cgttaacatt atttccgctg ctgccactat tcgaaaccaa 1200
 ggttaacttt ctgctgattc tgaagcaaaa gatataaagg gcaaatattg tctttccgcc 1260
 aaagagggtg aagcggaaat tggcggtgta atttccgctc aaaaatcacga agctaaaagg 1320
 gccagctga tgattacagg cgataaaggc acattaaaaa cagggtcagt tatcgacctt 1380
 tcaggtaaa aggggggaga aacttacctt ggcggtgacg agcgcggcga aggtaaaaac 1440
 gccattcaat tagcaaaaga aacctcttta gaaaaaggct caaccatcaa tgatcaggc 1500
 aaagaaaaag cgcgacgcgc tattgtgtgg gcgcatattg cgtttaattga cggcaatatt 1560
 aacgctcaag gtatgggtga tatcgctaaa accggtgggt ttgtggagac atcggggcat 1620
 tatttatcca ttgacagcaa tgcaattgtt aaaaacaaaag agtgggtgct agacctgat 1680
 gatgtatacaa ttgaagcga agaccccttt gcgaataata cgtatgaattc 1740
 ccaacaggca ccggtgaagc aagcgacctt aaaaaaata gcgaactcaa aacaacgcta 1800
 accaatacaa ctatttcaaa ttatctgaaa aacgctgcaa caatgaatc aacggcatac 1860
 agaaaactta ccgttaatag ctcaatcaac atcggaagca actccactt aattctccat 1920
 agtaaaaggc agcgtggcgg agcggttcag attgatggag atattacttc aaaggcgga 1980
 aatttaacca ttattctcgg cggatgggtt gatgttata aaaaattacc gcttgatcag 2040
 ggttttttaa atattaccgc gccttcgcga gcttttgagc gtggaaataa caaagcacgc 2100
 gcgcggcga atgctaaaaa tgcgcgccag ggcaactgaa ccattacagg agagggaaaa 2160
 gatttcaggg ctaacaacgt atctttaaacc ggaacgggta aaggctgtaa tatcatttca 2220
 tcagtgaata atttaaccga caatcttagt gcacataata gaataataa gaataataa 2280
 attaaccaaa ctacagaaaa gaacacctcg tatgggcaaa ccagccatga ttccgactgg 2340
 aacgctaggt cctttaaact agagacaggc gcaaatttta cctttattaa atacatttca 2400
 agcaatagca aaggcttaac aacacagtat agaagctctg cagggggtgaa ttttaacggc 2460
 gtaaatggca acatgtcatt caatctcaaa gaaggaagca aagtttaatt caaatataa 2520
 ccaaacgaga acatgaacac aagcaaacct ttaccaatcc ggttttttag caatatcaca 2580
 gccactgggt ggggctctgt tttttttgat atatatctca accattctgg cagaggggct 2640
 gagttcaaaa tgaagtgaat taatatctct aacggcgcta attttacctt aaattcccat 2700
 gctcgcggcg atgacgcttt taaatcaaac aaagacttaa ccatataatgc aaccaattca 2760
 aatttcagcc tcagacagac gaagatgat ttttatgacg ttatgacgc tagtccatc 2820
 aattcaacct acaacatcac cattctgggc ggtaatgtca cccttgggtg acaaaactca 2880


```

agcagcagca ttacggggaa tattactatc gagaagaacg caaatgttac gctagaagcc 2940
aataacgccc ctaatacagca aaacataaagg gatagagtta taaactctgg cagcttgctc 3000
tcagaagaagc gtttaagttt aactggcgaa aatgcagata ttaaggcaca tctcactatt 3060
tcagaagaagc ccacttttaa aggaagaact agagataacc taaatatcac cggaattttt 3120
accaataatg gcactgccga aattaatata acacaaggag tggtaaaact tggcaatgtt 3180
accaatgatg gtgattttaa cattaccact cactgtaaac gcaaccaaac aagcatcatc 3240
ggcgaggata taatcaacaa aaaaggaagc ttaaatatta cagacagtaa taatgatgct 3300
gaaatccaaa ttggcgcgaa tatctcgcaa aaagaaggca acctcacgtt ttcttccgat 3360
aaaatataa tcaccaaaca gataacaatc aaaaagggta ttgatggaga ggactctagt 3420
tcagatggca caagttaatgc caacctaaact attaaaacca aagaattgaa attgcagaaa 3480
gacctaaagta ttccaggttt caataaagca gagattacag ccaaaagatgg tagagattta 3540
actattggca acagtaatga cggtaaacgc ggtgccgaag ggtcacaaatg tgacactaaa tagcaaatg 3600
aatgttaaag attcaaaaat ctctgctgac ggtcacaaatg acgaatagcg ccaaaacagt aacttttaac 3660
aaaaactcta gcagcaatgg cggacgtgaa agcaatagcg ccaaaacagt aacttttaac 3720
attactgcga aaaaatgtaga agtaaacaaa gatattact ctctcaaaac agtaaatatc 3780
accgcgtcgg aaaaaggttac caccacagca ggctcgacca ttaacgcgcaac aaatggcaaaa 3840
gcaagttata caacaaaac aggtgatatc agcgctacga tttccggtaa caccgtaagt 3900
gttagcgcca ctggtgattt aaccactaaa tccggctcaa aaattgaagc gaaatcggtg 3960
gaggctaatg taacaagtgc aacaggtaca attggcggtt caatttccgg taatcggta 4020
aatgttaacg caaacgctgg cgatttaaca gttgggaatg gcgcagaaat taatgcgaca 4080
gaaggagctg caaccttaac cgcaacaggg aataccttga ctactgaagc cggttctagc 4140
atcacttcaa ctaagggtca ggtagacctc ttggctcaga atggtagcat cgcaggaagc 4200
attaatgctg ctaatgtgac attaaatatt acaggcactt taaccacgct ggcaggtcgc 4260
gatattaaag caaccagcgg cacttggtt attaacgcaa aagatgctaa gctaaatggg 4320
gtgctgagc gtgatatgac agaagtgaat gcagtcaacg caagcggttc tggtagtggt 4380
actggcgca cctcaagcag tgtgaatatt cactgggatt taaacacagt aaatgggta 4440
aatatcattt cgaaagatgg tagaaacact gtgcgcttaa gaggcagga aatgaggtg 4500
aaatatatcc agccaggtgt agcaagtgtt gaagaagtaa tgaagcgaa acgcgtctct 4560
gaaaaagtaa aagatttatc tgatgaagaa agagaacata tagctaaact tgggtgaagt 4620
gctgtacggt ttgttgagcc aaataatata attacagta atacacaaaa tgaattttaca 4680
accagacgtt caagtcaagt gataatttct gaaggttaag cgtgtttctc aagtgttaat 4740
ggcgacagag tatgtaccaa tttgtgctgac gatggacgc cgtagtcagt aattgacaag 4800
gtagatttca tctcgcaatg aagtcatttt atttctgtt tatttactgt gtgggttaaa 4860
gttcagtagc ggctttacc atcttgtaaa aaattacgga gaatacaata aagttatttt 4920
aacaggttat tattatg

```

<210> 4
 <211> 1477
 <212> PRT
 <213> Haemophilus influenzae

```

<400> 4
Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu
  1             5             10             15
Val Ala Val Ser Glu Leu Ala Arg Gly Cys Asp His Ser Thr Glu Lys
      20             25             30
Gly Ser Glu Lys Pro Ala Arg Met Lys Val Arg His Leu Ala Leu Lys
      35             40             45
Pro Leu Ser Ala Met Leu Leu Ser Leu Gly Val Thr Ser Ile Pro Gln
      50             55             60
Ser Val Leu Ala Ser Gly Leu Gln Gly Met Asp Val Val His Gly Thr
      65             70             75             80

```

Ala Thr Met Gln Val Asp Gly Asn Lys Thr Ile Ile Arg Asn Ser Val
85 90 95

Asp Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met
100 105 110

Val Gln Phe Leu Gln Glu Asn Asn Asn Ser Ala Val Phe Asn Arg Val
115 120 125

Thr Ser Asn Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly
130 135 140

Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
145 150 155 160

Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
165 170 175

Glu Asn Ile Lys Ala Arg Asn Phe Thr Phe Glu Gln Thr Lys Asp Lys
180 185 190

Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
195 200 205

Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
210 215 220

Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
225 230 235 240

Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
245 250 255

Glu Asn Glu Ala Val Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn
260 265 270

Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Gln Gly Lys Leu Ser Ala
275 280 285

Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
290 295 300

Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
305 310 315 320

Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
325 330 335

Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
340 345 350

Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala
355 360 365

Lys Lys Thr Ser Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
370 375 380

Glu Lys Gly Gly Phe Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
385 390 395 400

Gly Asn Ile Asn Ala Gln Gly Ser Gly Asp Ile Ala Lys Thr Gly Gly
 405 410 415
 Phe Val Glu Thr Ser Gly His Asp Leu Phe Ile Lys Asp Asn Ala Ile
 420 425 430
 Val Asp Ala Lys Glu Trp Leu Leu Asp Phe Asp Asn Val Ser Ile Asn
 435 440 445
 Ala Glu Asp Pro Leu Phe Asn Asn Thr Gly Ile Asn Asp Glu Phe Pro
 450 455 460
 Thr Gly Thr Gly Glu Ala Ser Asp Pro Lys Lys Asn Ser Glu Leu Lys
 465 470 475 480
 Thr Thr Leu Thr Asn Thr Thr Ile Ser Asn Tyr Leu Lys Asn Ala Trp
 485 490 495
 Thr Met Asn Ile Thr Ala Ser Arg Lys Leu Thr Val Asn Ser Ser Ile
 500 505 510
 Asn Ile Gly Ser Asn Ser His Leu Ile Leu His Ser Lys Gly Gln Arg
 515 520 525
 Gly Gly Gly Val Gln Ile Asp Gly Asp Ile Thr Ser Lys Gly Gly Asn
 530 535 540
 Leu Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn Ile Thr
 545 550 555 560
 Leu Asp Gln Gly Phe Leu Asn Ile Thr Ala Ala Ser Val Ala Phe Glu
 565 570 575
 Gly Gly Asn Asn Lys Ala Arg Asp Ala Ala Asn Ala Lys Ile Val Ala
 580 585 590
 Gln Gly Thr Val Thr Ile Thr Gly Glu Gly Lys Asp Phe Arg Ala Asn
 595 600 605
 Asn Val Ser Leu Asn Gly Thr Gly Lys Gly Leu Asn Ile Ile Ser Ser
 610 615 620
 Val Asn Asn Leu Thr His Asn Leu Ser Gly Thr Ile Asn Ile Ser Gly
 625 630 635 640
 Asn Ile Thr Ile Asn Gln Thr Thr Arg Lys Asn Thr Ser Tyr Trp Gln
 645 650 655
 Thr Ser His Asp Ser His Trp Asn Val Ser Ala Leu Asn Leu Glu Thr
 660 665 670
 Gly Ala Asn Phe Thr Phe Ile Lys Tyr Ile Ser Ser Asn Ser Lys Gly
 675 680 685
 Leu Thr Thr Gln Tyr Arg Ser Ser Ala Gly Val Asn Phe Asn Gly Val
 690 695 700

Asn Gly Asn Met Ser Phe Asn Leu Lys Glu Gly Ala Lys Val Asn Phe
 705 710 715 720
 Lys Leu Lys Pro Asn Glu Asn Met Asn Thr Ser Lys Pro Leu Pro Ile
 725 730 735
 Arg Phe Leu Ala Asn Ile Thr Ala Thr Gly Gly Gly Ser Val Phe Phe
 740 745 750
 Asp Ile Tyr Ala Asn His Ser Gly Arg Gly Ala Glu Leu Lys Met Ser
 755 760 765
 Glu Ile Asn Ile Ser Asn Gly Ala Asn Phe Thr Leu Asn Ser His Val
 770 775 780
 Arg Gly Asp Asp Ala Phe Lys Ile Asn Lys Asp Leu Thr Ile Asn Ala
 785 790 795 800
 Thr Asn Ser Asn Phe Ser Leu Arg Gln Thr Lys Asp Asp Phe Tyr Asp
 805 810 815
 Gly Tyr Ala Arg Asn Ala Ile Asn Ser Thr Tyr Asn Ile Ser Ile Leu
 820 825 830
 Gly Gly Asn Val Thr Leu Gly Gly Gln Asn Ser Ser Ser Ile Thr
 835 840 845
 Gly Asn Ile Thr Ile Glu Lys Ala Ala Asn Val Thr Leu Glu Ala Asn
 850 855 860
 Asn Ala Pro Asn Gln Gln Asn Ile Arg Asp Arg Val Ile Lys Leu Gly
 865 870 875 880
 Ser Leu Leu Val Asn Gly Ser Leu Ser Leu Thr Gly Glu Asn Ala Asp
 885 890 895
 Ile Lys Gly Asn Leu Thr Ile Ser Glu Ser Ala Thr Phe Lys Gly Lys
 900 905 910
 Thr Arg Asp Thr Leu Asn Ile Thr Gly Asn Phe Thr Asn Asn Gly Thr
 915 920 925
 Ala Glu Ile Asn Ile Thr Gln Gly Val Val Lys Leu Gly Asn Val Thr
 930 935 940
 Asn Asp Gly Asp Leu Asn Ile Thr Thr His Ala Lys Arg Asn Gln Arg
 945 950 955 960
 Ser Ile Ile Gly Gly Asp Ile Ile Asn Lys Lys Gly Ser Leu Asn Ile
 965 970 975
 Thr Asp Ser Asn Asn Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser
 980 985 990
 Gln Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Ile Asn Ile Thr
 995 1000 1005
 Lys Gln Ile Thr Ile Lys Lys Gly Ile Asp Gly Glu Asp Ser Ser Ser
 1010 1015 1020

Asp Ala Thr Ser Asn Ala Asn Leu Thr Ile Lys Thr Lys Glu Leu Lys
 1025 1030 1035 1040
 Leu Thr Glu Asp Leu Ser Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr
 1045 1050 1055
 Ala Lys Asp Gly Arg Asp Leu Thr Ile Gly Asn Ser Asn Asp Gly Asn
 1060 1065 1070
 Ser Gly Ala Glu Ala Lys Thr Val Thr Phe Asn Asn Val Lys Asp Ser
 1075 1080 1085
 Lys Ile Ser Ala Asp Gly His Asn Val Thr Leu Asn Ser Lys Val Lys
 1090 1095 1100
 Thr Ser Ser Ser Asn Gly Gly Arg Glu Ser Asn Ser Asp Asn Asp Thr
 1105 1110 1115 1120
 Gly Leu Thr Ile Thr Ala Lys Asn Val Glu Val Asn Lys Asp Ile Thr
 1125 1130 1135
 Ser Leu Lys Thr Val Asn Ile Thr Ala Ser Glu Lys Val Thr Thr Thr
 1140 1145 1150
 Ala Gly Ser Thr Ile Asn Ala Thr Asn Gly Lys Ala Ser Ile Thr Thr
 1155 1160 1165
 Lys Thr Gly Asp Ile Ser Gly Thr Ile Ser Gly Asn Thr Val Ser Val
 1170 1175 1180
 Ser Ala Thr Val Asp Leu Thr Thr Lys Ser Gly Ser Lys Ile Glu Ala
 1185 1190 1195 1200
 Lys Ser Gly Glu Ala Asn Val Thr Ser Ala Thr Gly Thr Ile Gly Gly
 1205 1210 1215
 Thr Ile Ser Gly Asn Thr Val Asn Val Thr Ala Asn Ala Gly Asp Leu
 1220 1225 1230
 Thr Val Gly Asn Gly Ala Glu Ile Asn Ala Thr Glu Gly Ala Ala Thr
 1235 1240 1245
 Leu Thr Ala Thr Gly Asn Thr Leu Thr Thr Glu Ala Gly Ser Ser Ile
 1250 1255 1260
 Thr Ser Thr Lys Gly Gln Val Asp Leu Leu Ala Gln Asn Gly Ser Ile
 1265 1270 1275 1280
 Ala Gly Ser Ile Asn Ala Ala Asn Val Thr Leu Asn Thr Thr Gly Thr
 1285 1290 1295
 Leu Thr Thr Val Ala Gly Ser Asp Ile Lys Ala Thr Ser Gly Thr Leu
 1300 1305 1310
 Val Ile Asn Ala Lys Asp Ala Lys Leu Asn Gly Asp Ala Ser Gly Asp
 1315 1320 1325

Ser Thr Glu Val Asn Ala Val Asn Ala Ser Gly Ser Gly Ser Val Thr
 1330 1335 1340

Ala Ala Thr Ser Ser Ser Val Asn Ile Thr Gly Asp Leu Asn Thr Val
 1345 1350 1355 1360

Asn Gly Leu Asn Ile Ile Ser Lys Asp Gly Arg Asn Thr Val Arg Leu
 1365 1370 1375

Arg Gly Lys Glu Ile Glu Val Lys Tyr Ile Gln Pro Gly Val Ala Ser
 1380 1385 1390

Val Glu Glu Val Ile Glu Ala Lys Arg Val Leu Glu Lys Val Lys Asp
 1395 1400 1405

Leu Ser Asp Glu Glu Arg Glu Thr Leu Ala Lys Leu Gly Val Ser Ala
 1410 1415 1420

Val Arg Phe Val Glu Pro Asn Asn Thr Ile Thr Val Asn Thr Gln Asn
 1425 1430 1435 1440

Glu Phe Thr Thr Arg Pro Ser Ser Gln Val Ile Ile Ser Glu Gly Lys
 1445 1450 1455

Ala Cys Phe Ser Ser Gly Asn Gly Ala Arg Val Cys Thr Asn Val Ala
 1460 1465 1470

Asp Asp Gly Gln Pro
 1475

<210> 5
 <211> 9171
 <212> DNA
 <213> Haemophilus influenzae

<400> 5
 acagcgctct cttaataacta gtacaaaccc acaataaaaat atgacaaaaca acaattacaa 60
 cacccttttt gcagctctata tgcaaatatt ttaaaaaata gtataaatcc gccatataaa 120
 atggatataat ctttcatctt tcattcttca tctttcatct ttcatcttcc atctttcatc 180
 ttctcatcttt catctttcat ctttcatctt tcattcttca tctttcatct ttcatcttcc 240
 acatgaaatg atgaaccgag ggaagggagg gaggggcaag aatgaagagg gagctgaacg 300
 aacgcaaatg ataaagtaat ttaattgttc aactaacctt agggagaaaa atgaacaaga 360
 tatatcgtct caaattcagc aaacgcctga atgccttggg tgctgtgtct gaattggcac 420
 ggggtgttga ccattccaca gaaaaaggca gcgaaaaacc tgctcgcatg aaagtgcgtc 480
 acttagcggt aaagccaact tccgctatgt tactatcttt aggtgtaaca tctattccac 540
 aatctgtttt agcaagcggc ttacaaggaa tggatgtagt acacggcaca gccactatgc 600
 aagtagatgg taataaaaacc attatccgca acagtggtga cgctatcatt aattggaaac 660
 aatttaacat cgacccaaaat gaaatgggtgc agtttttaca agaaaaaac aactccgcgc 720
 tattcaaccg tggtacatct aacccaaatct cccaattaaa agggatttta gattctaacg 780
 gacagctctt tttaatcaac ccaaatggta tcacaatagg taaagacgca attattaaca 840
 ctaatggctt tacggtctct acgctagaca tttctaagca aaacatcaag gcgcgtaatt 900
 tcaccttcga cgaacccaaa gataaagcgc tgcgtgaaat tgtgaatcac ggtttaattt 960
 ctgctcggtaa agacggcagc gtaaatctta ttggtggcaa agtgaaaaac gaggtgtgtg 1020
 ttagcgtaaa tgggtggcagc attcttttac tcgcagggca aaaaatcacc atcagcgata 1080
 taataaaaccc aaccattact tacagcattg ccgcgcctga aaatgaagcg gtcaatctcg 1140
 cggatatttt tgcgaaggcg ggttaacatta atgtccgtgc tgccactatt cgaacccaag 1200
 ctttccgcga aagaggggtg agcggaaatt ggcggtgtaa ttccgcctca aaatcagcaa 1260
 gctaaggcgc gcaagctgat gattacagcg gataaagta cattaaaaac aggtgcagctt 1320

atcgaccttt	caggtaaaga	agggggagaa	acttaccttg	gcggtagaca	gcgcggcgaa	1380
ggtaaaaatc	gacttcaatt	agcaaaagaa	acctctttag	aaaaaggctc	aaccatcaat	1440
gtatcaggca	aagaaaaaag	cggacgcgct	attgtgtggg	gcgatattgc	gttaattgac	1500
ggcaatatta	acgctcaaa	tagtggtgat	atcgctaaaa	ccggtggttt	tgtggagacg	1560
tgccggcgat	attttatcat	caaagacaat	gcaattgttg	acgccaaaga	gtggtgtgtta	1620
gacccggata	atgtatctat	taatgcagaa	acagcaggac	gcagcaatca	ttcagaagac	1680
gagatgaata	cgggatccgg	gaatagtgcc	agcaccceaa	aacgaaacaa	agaaaagaca	1740
acattaaaca	acacaactct	tgagagtata	ctaaaaaaag	gtacctttgt	taacatcact	1800
cgctaataca	gcacttatgt	caatagctcc	attaatttat	ccaatggcag	cttaactctt	1860
tggagtgtgt	gtcggagcgg	tggcggcggt	gagatataca	acgatattac	acccggtgat	1920
gataccagag	gtgcaaacct	acaattttac	tcaggcgctg	gggttgatgt	tcataaaaaa	1980
atctcactcg	gggcgaag	taacataaac	attacagcta	aacaagatat	cgcttttgag	2040
aaaggaagca	accaagtcac	tacaggtcaa	gggactatta	cctcaggcaa	tcaaaaaggt	2100
tttagattta	ataatgtctc	tctaaacggc	actggcagcg	gactgcaatt	caccactaaa	2160
agaaccaata	atacgcgtat	cacaaaataa	tttgaaggga	ctttaaatat	ttcagggaaa	2220
gtgaacatct	caatggtttt	acctaaaaat	gaaagtggat	atgataaatt	caaaaggacg	2280
acttactgga	atttaacctc	gaaagtggat	atgataaatt	caaaaggacg	ctccactatt	2340
gactccagag	gaagcgatag	tgcaggcaca	cttaccacag	cttataaatt	aaacgggata	2400
tcatttaaca	aagacactac	ctttaatggt	gaacgaaatg	caagagtcaa	ctttgacata	2460
aagggcccaa	tagggataaa	taagtattct	agtttgaatt	acgcattcatt	taatggaaac	2520
atttcagttt	cggagggggg	gagttgtgat	ttcacacttc	tcgcctcact	ctctaacgct	2580
caaaccctcg	gtgtagttat	aaattctaaa	tactttaatg	tttcaacagg	gtcaagttta	2640
agatttaaaa	cttcaggctc	aacaaaaact	ggcttctcaa	tagaagaaga	tttaacttta	2700
aatgccacgc	gaggcaacat	aacacttttg	caagttgaag	gcaccgatgg	aatgatttgt	2760
aaaggcattg	tagccaaaaa	aaacataacc	tttgaaggag	gtaagatgag	gttttggcttc	2820
agggaaagcc	taacagaaat	cgaaggcaat	gttactacta	atacaacgcg	taacgtcaat	2880
cttatcggtt	cggattttga	caaccatcaa	aaactcttaa	ctattaaaaa	agatgtcatc	2940
attaatagcg	gcaacctttac	cgcgtggagg	aatattgttc	atatagccgg	aaactctaac	3000
gttgaagata	acgctaattt	caaaagctatc	acaaatttca	cttttaattg	agggcgcttg	3060
tttgacaaca	aaggcaattc	aaatatttcc	attgccaaag	gaggggctcg	ctttaaagac	3120
attgataatt	acgaagaatt	aagcatcacc	accaactcca	gctccactta	ccgcactatt	3180
ataaaggcga	atataaccac	taaaaaacgt	gatttaaat	ttacgaacga	aggtatgtat	3240
actgaaatgc	aaattggcgg	cgatgtctcg	caaaaagaag	gtaatctcac	gatttctctc	3300
gacaaaatca	atattaccac	acagataaca	atcaaggcag	gtgtgtatgg	ggagaattcc	3360
gattcagacg	cgacaacac	tgccaattca	accattaaaa	ccaaagaatt	gaaattcaacg	3420
caagacctaa	atatttcagg	tttcaataaa	gcagagatta	cagctaaaga	tggtatgtat	3480
ttaactattg	gtacacccaa	tagtgctgat	ggtactaatg	ccaaaaaagt	aaactttaac	3540
caggtttaaag	atacaaaaat	ctctgctgac	ggtcacaagg	tgacactaca	cagcaaaagt	3600
gaaacctccg	tagtataata	caacactgaa	gtatagcagt	acaataatgc	cggtcttaact	3660
atcgatgcaa	aaaatgtaac	agtaaaccaac	aatattactt	ctcaaaaagc	agtgagcatc	3720
tctgcgacaa	gtggagaaat	taccactaaa	acaggtacaa	ccattaacgc	aaccacttgt	3780
aacgtggaga	taaccggtca	aacaggtagt	atcctagggt	gaattgagtc	cagctctggc	3840
tctgttaacac	ttactgcac	cgaaggcgct	cttgcgttaa	gcaatatttc	gggcaaacac	3900
gttactgtta	gtccaaatag	cgggtgcatta	accacttttg	caggctctac	aattaaaggga	3960
accgagagtg	taaccacttc	aagtcaatca	ggcgatctgc	gcggtacgat	ttctgtggcg	4020
acagttagagg	ttaagacac	cgaagtttta	ccacttcaaa	ccaatcaaaa	aattaaagca	4080
acaaacaggc	aggctcaagt	aacaagtgtca	acaggtacaa	ttggttggtat	gatttccggt	4140
aatcgcgttaa	agtttaacgc	aaacgcgtgc	gatttaacag	ttgggaatgt	ccgagaagtt	4200
aatcgcagac	aaggagctgc	aaccttaact	acatcatcgg	gcaaattaac	taccggaagt	4260
agttcacaca	ttacttcagc	caagggtcag	gtaaatcttt	cagctcagga	tggtagcgtt	4320
gcaggaagta	ttaatgccgc	caatgtgaca	ctaaatacta	caggcacttt	aactaccgtg	4380
aaaggttcaa	acttaattgc	aaccagcggt	accttgggtta	ttaacgcaaa	agacgctgag	4440
ctaaatggcg	cagcatctgg	taaccacaca	gtgttaaatg	caaccaacgc	aaatggctcc	4500
ggcagcgtaa	tccgcgcaac	ctcaagcaga	gtgaacatca	ctggggattt	aatcacataa	4560
aattgatttaa	atatcatttc	aaaaaacggg	ataaacctcg	tactgttaaa	agggcgttaa	4620
attgatgtga	aatcaattca	acggggtata	gcaagcgtag	atgaagtaat	tgaagcgaaa	4680
cgcatctctg	gaagaggtaaa	agatttatct	gatgaagaaa	gagaagcggt	agctaaactt	4740
ggcgtaagtg	ctgtacgttt	tattgagcca	aataatacaa	ttacagtcga	tacacaaat	4800
gaatttgcaa	ccagaccatt	aagtgaata	gtgattctcg	aaggcaggcg	gtgtttctca	4860

aacagtgatg	gcgcgacggt	gtgcgttaat	atcgctgata	acgggcggta	gcggtcagta	4920
atgtgacaag	tagattttcat	cctgcaatga	agtcatttta	ttttcgttatt	atttactgtg	4980
tgggttaaa	ttcagtgacg	gctttaccga	tcttgtaaaa	aattacgggag	aatacaataaa	5040
agtattttta	acaggtttatt	attatgaaaa	atataaaaa	cagattaaaa	ctcagtgcaa	5100
tatcagctatt	gcttggcctg	gctctttcat	cattgtatgc	agaagaagcg	tttttagtaa	5160
aaggctttca	gttatctggg	gcacttgaaa	ctttaagtga	agacgcccaa	ctgctctgtag	5220
caaaattctt	attcaaaatc	caaggctcgc	aaactttaac	aaacctaaaa	acgacgcagc	5280
ttgaattaca	ggctgtgcga	gataaagatt	agccaaataa	grrtgatgtg	atatgtccac	5340
aaacaaacat	tacggatggc	aatatattgt	ttgagctagt	ctcgaaatca	gcgcgacgaa	5400
gccaaatttt	ttataaggcg	agccagggtt	ctagtgaaga	aaatactcgt	cgtagctcgc	5460
catctttgaa	acaaagaaaa	gtgtatgaag	atggtcgta	gtggttcgat	ttgcgtgaat	5520
tcaatatggc	aaaaaagaaa	ccacttaaa	tcactcgcgt	gcattacgag	ttaaacctta	5580
aaaacaaaa	ctctgatttg	gtagttgcag	gtttttcgcc	ttttggcaaa	acgcgtagct	5640
ttttttctca	tgataatttc	ggcgcaagg	agtttaacta	tcaacgtgta	agtcgtaggt	5700
ttgtaaatgc	caatttgacc	ggacatgatg	atgtatata	tctaaacgta	ttgaccaatg	5760
taaaagcacc	atcaaaatct	tatgcggtag	gcataaggata	tactttatccg	tttatgata	5820
aacaccaatt	cttaagtctt	tataccagca	tgagttagtc	tgattctaat	gatatcgacg	5880
gctttaccaa	tcggattaat	cgtaaatatt	caaaaggta	atctatctct	gcgaatctga	5940
aatggagtta	ttactctccg	acatttaacc	ttggaatgga	agaccagttt	aaaatttaatt	6000
taggctacaa	ctaccgcat	atataacaaa	catccgagtt	aaacacctgt	gggtgcaagca	6060
agaaaaaatt	tcagtatca	ggcgtaagtg	caggcatgtg	tgacatactc	caattttacc	6120
ctaaaaacaat	ctttaatatt	gatttaactc	atcattatta	cgcgagtaaa	ttaccaggct	6180
cttttggaa	ggagcgcatt	ggcgaaacat	ttaatcgcag	ctatcacatt	agcacagcca	6240
gttttagggt	gagtgcaagag	tttgctcaag	gttggcattt	tagcagtgta	ttatcggggt	6300
agttttactc	acaagatata	agtagcatag	atttattctc	tgtaaacaggt	actttatggc	6360
tcagaggctt	taaatcaggc	gtgtgcaagt	gtgagcgcgt	tcttgatgtg	cgtaatgaa	6420
taagtatgcc	aaaatcaccc	cgctttcaaa	tcagccctta	tgcgttttat	gatgcaggtc	6480
agtttcggtta	taatatgcgaa	aatgtcaaaa	cttaccggca	agatatgcac	acggatctct	6540
ctgcggggtt	agggcattaaa	acctctctca	cacaaaaact	aagcttagat	gctttttgtg	6600
ctcgtcgctt	tgcaaatgcc	aatagtgaca	ttttgaatgg	caacaaaaaa	cgacaaagct	6660
cacctacaac	ctctcggggt	agattaaact	atgtttctca	acctgaaat	ttaatcaact	6720
ggtaagcggt	ccgcctacca	gtttataact	atatgcttta	ccgcgcaatt	tacagctcat	6780
acgcaacctc	gttttctatc	ttatatatca	aacaaactaa	gcaaaaccaag	caaaaccaagc	6840
aaaccaagca	aaccaagcaa	accaagcaaa	ccaagcaaac	caagcaaaac	aagcaaaacca	6900
agcaaaacaa	gcaaaaccaag	caaaaccaagc	aaaccaagga	atgctaaaaa	acaatttata	6960
tgataaacta	aaacatactc	cataccatgg	caatacaagg	gattttaata	tatgacaaaa	7020
gaaaaattac	aaagtgttcc	acaaaatagc	acgccttcac	ttgtagaact	aaacaacgac	7080
caaaactccc	tgcaaaatct	taaaaacca	cccaaaccca	acctattacg	cttggaacaa	7140
catgtgcgca	aaaaagatta	tgagctttgt	tgccgcgaat	taatggcgat	tttggaaaaa	7200
atggagacta	attttggagg	cgctcacgat	attgaaattt	acgcacctgc	tcagctggga	7260
ctcttaccgc	aaaaactact	aattcatttt	gccactcgct	tcgctaagtc	aattacaaca	7320
tctttttctc	accccgcaatt	ggcaatttcc	gaagaagggg	ctataaagat	gattagctgt	7380
caacgctggt	tgacgctgat	ttttgctctt	tcccctacg	ttaacgcaga	ccatattctc	7440
ataataatata	atataacccc	agattccgaa	gttggtcttc	atttagcaac	agacaactct	7500
tctatttgcta	aattctgtat	tttttaacta	ccggaatcca	atgtcaatat	gagtttagat	7560
cggttatggg	cagggaatca	acaactttgt	gcttcatgtt	gttttgcgtt	gcagctttca	7620
cgttttatgt	gtactgcata	tgcttttcat	aaaaagacgg	tggttttaca	gtggtttcct	7680
aaaaaacctc	ccgaaattgc	taatttagat	gaattgcctg	caaatatcct	ctatgatgta	7740
tatatgtca	gcagttatga	tttagcaaaa	aacaagcacg	atgttaagcg	ttcattaaac	7800
gaacttgtcc	gcaagcatat	cctcacgcga	ggatgggca	acgcgtacct	ttacacctta	7860
ggtaaaaagg	acgggcaaac	tgtgatgatg	gtactgctgt	aacattttaa	ttcgggacat	7920
tcgattttatc	gcacgcattc	aacttcaatg	attgctgtct	gagaaaaatt	ctatttagtc	7980
ggcttagggc	ataggggcgt	tgataacata	ggtcgagaag	gttttgacga	gttctttgaa	8040
atcagtagca	ataataataa	ggagagactg	ttttttatcc	gtaaacagtg	cgaacctttc	8100
caaccgcgag	tggttcttat	gccaaagcatt	ggcattggata	ttacctcagat	ttttgtgagc	8160
aacctccggc	ttgcccctat	tcaagctgta	gctttggggtc	atccctgccac	tacgcatctc	8220
gaatttatgt	attatgtcat	cgtagaagat	gattatgtgg	cgagtgaaag	ttgttttagc	8280
gaacaccttt	tacgctttac	caaagatgcc	ctacctttatg	taccatctgc	actcggccca	8340
caaaaagtgg	attatgtact	cagggaaaaac	cctgaagtga	tcaatatcgg	tattgcgcgt	8400


```

accacaatga aattaaaccc tgaatttttg ctaacattgc aagaaatcag agataaagct 8460
aaagtcacaa tacatttttca tttcgacatt ggacaatcaa caggcttgac acacccttat 8520
gtcaaatggg ttatcgaaag ctatttaggt gacgatgcca ctgcacatcc ccacgcacct 8580
tatcacgatt atctggcaat attcggtgat tgcgatatgc tactaaatcc gtttcccttc 8640
ggtaatacta accgcataat tgatatgggt acattagggt tagttgggtg atgcaaaacg 8700
ggggatgaag tacatgaaca tattgatgaa ggtctgttta aacgcttagg actaccagaa 8760
tggctgatag ccgacacacg agaaacatat attgaatgtg atcatagaaa acaacgcgtt accaaaaagc 8820
catcaagaac gccttgaact ccgtcggttac atcatagaaa agaaaaacaa tgaatggaag 8940
tttacaggcg accctcgctcc attggggcaaa atactgctta agaaaaacaa atgtttcaaa 9000
cggaatactg tgaagtaaaa ataacggttt tttaaagtaa aagtcgggtt tcccgcgcg 9060
gcgttttaaa aacctctcaa aaatcaacgc cacttttacc ttataacgc tcccgcgcg 9120
tgacagttta tctctttctt aaaataccca taaaattgtg gcaatagttg gtaatacaaa 9120
ttcaattggt gatacggcaa actaaagacg gcgcgttctt cggcagtcac c 9171

```

<210> 6

<211> 9323

<212> DNA

<213> Haemophilus influenzae

<400> 6

```

cgccacttca attttggatt gttgaaattc aactaaccaa aaagtgcggt taaaatctgt 60
ggagaaaaat ggtttgtagt aagaacgagg taattgttca aaaggataaa gctctcttaa 120
ttgggcattg ttggcgcttt ctttttcggt taatagtaaa ttatatctg gacgactatg 180
caatccacca acaactttac cggtggtttt aagcgtaaat gtaagtctct gctctctctg 240
gcgaatacgt aatccccatt ttgttttagc agaaaaatga tcgggataat cataataggt 300
gttgcccaaa aataaatttt gatgttctaa aatcataaat ttgcaagat attgtggcaa 360
ttcaatacct attttgtggc aaatcgccaa ttttaattca atttctgtga gcataatatt 420
tcccactcaa atcaactggg taaatatata agataataaa aataaatcaa gatttttgtg 480
atgacaaaca acaattacaa cactttttt gcagctcata tgcaaatatt ttaaaaaaat 540
agtataaatc cgccatataa aatggataaa tctttcatct ttcattcttc atctttcatc 600
tttcatcttt catctttcat ctttcatctt caatgaaat gatgaaccga ggggaaggag ggaggggcaa 720
atctttcatc attttcatc cactgaaat gaataagtaa ttaattgtt caactaacct 780
gaagtgaagag ggagctgaac gaacgcaaat gataaagtaa ttaattgtt caactaacct 780
tagggagaaaa tatgaacaag atatatcgtc tcaaaattcg caaacgcctg aatgctttgg 840
ttgctgtgtc tgaattggca cgggggtgtg accattccac agaaaaaggc agcgaaaaac 900
ctgctcgcat gaaagtcgct cacttagcgt taaagccaat ttcgcctatg ttactattct 960
taggtgttaac atctattcca caatctgttt tagcaacgag caatttaaca tcgacaaaaa 1020
tgaanaagtg cagtttttac aagaaaaaca cattatccg cattatccg acacgtgttg 1080
acgctatcat taattggaaa caatttaaca tcgacaaaaa tgaatgggtg cagtttttac 1140
aagaaaaaca caactccgcc gattttacac tgacaaaaa ttaccaaatc tcccaataaa 1200
aagggatttt aattattaac actaatggct ttacggcttc tacgctagac atttcaacg 1320
gtaagagcgc aaaaactcaa gcgcgctaatt ttcaccttcg agcaaaccaa agataaaagc ctgcgtgaaa 1380
tttgtaatca cggtttaatt actgtcggta aagacggcag tgtaaatctt atttggtgga 1440
aagtgaaaaa cgagggtgtg attagcgtaa attggtggcg catttcttta ctgcagggcg 1500
aaaaaatcac catcagcgat ataataaac caaccattac ttacagcatt gccgcgctg 1560
aaaaagaagc ggtcaatctg ggcgatat ttgcccagg cggttaacat aatgtccgtg 1620
ctgccactat tcgaatacaa ggtataactt ctgctgattc tgtaagcaaa gataaaagcg 1680
gcaatattgt tctttccgcc aaagagggtg aagcggaaat tggcgtgtga atttccgct 1740
aaaaatcaga atctaaaggc ggcaagctga tgataaagtc cgataaagtc acattaaaaa 1800
cagggtgcagt tatcgacctt tcaggtaaag aagggggaga aacttacctt ggcggtagcg 1860
agcgcggcga aggtaaaaac ggcattcaat tagcaaaaga accctcttta gaaaaaggct 1920
caacctatcaa tgatcaggc aaagaaaaag gcggacgcgc tattgtgtgg ggcgatattg 1980
cgttaattga gcgcaatatt aacgctcaag gtatgtgtga tatcgctaaa accggtggtt 2040
tttgtggagc atcggggcat tatttatcca ttgacagcaa tgcaattgtt aaaaaaaag 2100
agtgtgtctg agaccctgat gatgtaaaca ttgaagcga agacccccct cgcaataa 2160
ccggtataaa tgatgaattc ccaacaggca ccggtgaagc aagcgacctt aaaaaaata 2220
gcgaactcaa aacaacgcta accaatacaa ctatttcaaa ttatctgaaa aacgcctgga 2280

```

caatgaatat	aacggcatca	agaaaaacta	ccgttaatat	ctcaatcaac	atcggaagca	2340
actcccactt	aatctcccat	agtaaaaggtc	agcgtggcgg	aggcgttcag	atgtatggag	2400
atattacttc	taaaagcgga	aatttaacca	tttattctgg	cggatggggt	gaagttcata	2460
aaaaattacc	gcttgatcag	gggttttttaa	atattaccgc	cgcttcogta	gcttttgaag	2520
gtggaaataa	caaaagcagc	gacgcgggcaa	atgtcaaaat	tgtcgccag	ggcactgtaa	2580
ccattacagg	agagggaaaa	gatttcaggg	ctaaccaact	atctttaaac	ggacgggta	2640
aagggtcgaa	tatcatattca	tcagtgaaata	atttaaccca	caatcttagt	ggcacaatta	2700
acatatcgga	gaataataca	attaacccaa	ctacagcgaa	gaacacctgt	tattggcaaa	2760
ccagccatga	ttcgcaactgg	aacgtcagtg	ctcttaactt	agagacagcg	gcataatttt	2820
ccctttataa	atacatattca	agcaatagca	aaggctttaac	aacacagtat	agaagctctg	2880
cagggggtgaa	ttttaacggc	gtaaatggca	acatgtcatt	caatctcaaa	gaaggagcga	2940
aagtttaatt	caaatataaa	ccaaacgaga	acatgaacac	aagcaaacct	ttaccaattc	3000
gggttttagc	caatatcaca	gccactgggt	ggggctctgt	tttttttgat	atatatgcca	3060
accattctcg	cagagggggct	gagttaaaaa	tgagtgaat	taatatctct	aacggcgcta	3120
atttttacc	aaattcccat	gttcggcggt	atgacgcttt	taaaatcaac	aaagacttaa	3180
ccataaactg	aaccaattca	aatttcagcc	tcagacagac	gaaagatgat	ttttatgacg	3240
ggtagcagc	aagtcacatc	aattcaacct	acacacatc	cattctggcg	ggtaattgtca	3300
cccttggttg	acaaaactca	agcagcagca	ttacggggaa	tattactatc	gagaaagcga	3360
caaatgtttac	gctagaagcc	aataacgccc	ctaatcagca	aaacataaag	gatagagtta	3420
taaaaacttgg	cagcttgtctc	gttaatggga	gtttaagttt	aactggcgaa	aatgcagata	3480
ttaaaggcaa	tcttactatt	tcagaaaagc	ccacttttaa	aggaagaact	agagataccc	3540
taaatattcac	cgggcaatttt	accaataatg	gcactgcgca	aattaatata	acacaaaggag	3600
tggttaaaact	tggcgaatgtt	accaatgatg	gtgattttaa	cattaccact	cacgctaatac	3660
gcaaaactgaa	aaagcatcatc	ggcgagagata	taatacaaca	aaaaggaagc	ttaaaattata	3720
cagacagtaaa	taatatgtct	gaaatccaaa	ttggcgggcaa	tatctcgcaa	aaagaaagga	3780
acctcagcat	ttcttcogtat	aaaatttaata	tcaccaacaac	gatacaaatc	aaaaagggta	3840
ttgatggaga	ggactctagt	tcagatgcga	caagttaatgc	caacctaaact	atataaaacca	3900
agaagattgaa	attgcagcaa	gacctaaagta	tttcaggttt	caataaagca	gagattacag	3960
ccaaagatgg	tagagattta	actatttgcca	acagtaatga	cggttaacagc	gggtccgaag	4020
ccaaaaacagt	aacttttaac	aatgttaaag	attcaaaaat	ctctgctgac	gggtcaaatg	4080
tgacactaaa	tagcaaatgtg	aaaaacatcta	gcagcaatgg	cggaactgaa	agcaatagcg	4140
caaacgatac	cggtcttaact	attactgcga	aaaatgtaga	agtaaaacaaa	gatattctt	4200
ctctcaaaaac	agtaaaatatac	acccgctcgg	aaaaagttac	caccacagca	ggctcgacca	4260
ttaacgcac	aaatggcaaa	gcaagtatta	caaccaaaac	agggtatatac	agcggtacga	4320
tttcogtgtaa	cacgggtaaagt	gttagcgcga	ctgggtgatt	aaccactaaa	tcggctcaa	4380
aaattgaagc	gaaatcgggt	gaggctaagt	taacaagtgc	aacaggtaac	attggcggtta	4440
caattcccg	taatacggta	aatgltacgg	caaacgctgg	cgatttaacac	gttgggaagt	4500
gcgacagaaat	taatgcgaca	gaaggagctg	caaccttaac	cgcaacaggg	aataccttga	4560
ctactgaagc	cggtttctagc	atcacttcaa	ctaagggtga	ggtagacctc	ttggctcaga	4620
atggttagct	cgacaggaagc	attaatgtcat	ctaatgtgac	attaaatact	acaggcacct	4680
taaccacccgt	ggcaggtctg	gatattaaag	caaccagcgg	caccttggtt	attaaacgaa	4740
agaagtgtctaa	gctaagattgg	gatgcatacag	gtgatagta	acagtgtaac	cgagtaacg	4800
actggggatt	tggtagtgtg	actgcggcaa	cctcaagcag	tgtgaatac	actggggatt	4860
taaaacacagt	aaatgggttta	aatatcattt	cgaaagatgg	tagaaaacact	gtgcgcttaa	4920
gaggcaagga	aattgaggtg	aaatatatcc	agccaggtgt	agcaagtgtta	gaagaagtaa	4980
ttgaagcgaa	acgcgtctct	gaaaaagtaa	agaatttatac	tgatgaagaa	agagaagact	5040
tagctaaact	tggtgttaagt	gctgtacgtt	ttgttgagcc	aaataataca	attacagtca	5100
atacacaata	tgaattttaca	accagacccg	caagtcgaat	gataattttc	gaaggtgaag	5160
cggtgtttctc	aagtggttaat	ggcgacacag	tatgtaccaa	tgttgctgac	gatggacagc	5220
cgtagtcagt	aattgacaag	gtagatttca	tccctgcaat	aagtcatttt	attttcgtat	5280
tatttactgt	gtgggttttaa	gttcagtagc	gcctttacc	atcttgtaaa	aatcttcgga	5340
gaatacaata	aagttattttt	aacaggttat	tattatgaaa	aatataaaaa	cgagattcaa	5400
actcagtgca	atatcagtat	tgcttggcct	gcctttctca	tcatttgtatg	cgaagaagac	5460
gttttttagta	aaaggctttc	agttatctcg	tgactttgaa	acttttaagt	aaagacccca	5520
actcgtctga	gcataaactt	tatctaaata	ccaaggctcg	agataagatt	caaaccttaa	5580
aacagcacag	cttgaattac	aggctgtgct	agataagatt	gagccaaata	aatattgatg	5640
gatattgccc	caacaaacca	ttacggatgg	caatatcatg	tttgagctag	ctctgaaatt	5700
agccgcagaa	agccaaagtt	tttataagc	agccagaggt	tatagtgaa	aaaatatctg	5760
tcgtagctcg	ccatctttga	aacaaggaaa	agtgtatgaa	gatggctcgc	agtggttcga	5820

tttgcgtgaa	ttaaataatg	caaaagaaaa	cccgccttaag	gttaccocgtg	tacattacga	5880
actaaaccc	aaaaacaaaa	cctctaattt	gataattg	ggctcttcgc	cttttggtgaa	5940
aacgcgcagc	tttattcttt	atgataattt	cgccgcgaga	gagtttaact	accaacgtgt	6000
aagctctggg	tttgcgttaag	ccaatttaac	tggtcatgat	gatgtgttaa	ttatacaggt	6060
atgagttatg	ctgattctaa	tgatatcgac	ggcttaccaa	gtgcgattaa	tcgtaaaata	6120
tcaaaaggte	aatctatctc	tgcgaaatctg	aaatggagtt	attatctccc	aacattttaac	6180
cttgggcattg	aagaccaatt	taaaattaat	ttaggctcaa	actaccgcca	tattaatcaa	6240
acctcccggt	taaatcgctt	gggtgaaacg	aagaaaaaat	ttgcagtaac	aggcgtgaag	6300
gcaggcattg	atggacatat	ccaatttacc	cctaaaacaa	tctttaatat	tgattttaact	6360
catcattatt	acgcgcgtaa	attaccaggc	ctctttggaa	tgagcgcgat	tggtgaaaca	6420
tttaatcgca	gctatcacat	tagcacagcc	agtttagggt	tgagtcaga	gtttgctcaa	6480
ggttggcatt	tttagcagta	attatcaggt	caatttactc	tacaagatat	tagcagata	6540
gattttattct	ctgtaacagg	tacttatggc	gtcagaggct	ttaataacgg	cggtgcaagt	6600
ggtagagcgc	gtcttgatg	gcgtaatgaa	ttaaagtatg	caaaatacac	ccgcttccaa	6660
atcagccctt	atcggtttta	tgatgcaggt	cagttccggt	ataatagcga	aaatgctaaa	6720
acttaacggc	aagatatgca	cacggatccc	tctgcggggt	taggcattaa	aacctctccc	6780
acacaaaact	taagctcaga	tgcttttggt	gtctgctcgt	tcgcaaatgc	caaatgtgac	6840
ttattggaat	gcaacaaaaa	acgcacaagc	tcaactcaaa	cctctggggg	gagatataca	6900
ttcaggtttct	aaccctgaaa	tttaataaac	tggtaaagct	cgccctaccc	agttttataac	6960
tatatgtctt	accgcccaat	ttacagttca	taggcaaccc	tggtttttacc	cttatatact	7020
aaataaaaaa	ctgaagctga	gctaagcaaa	caaagcaaac	tcaagcaagc	caagtaatac	7080
taaaaaaaca	atttatatga	taaaactaa	tatactccat	gccatggcga	tacaagggat	7140
tttaataat	gacaaaaagaa	aaattgcaaa	acgctcccca	agatgcgacc	gctttacttg	7200
cgggaattaa	caacaatcaa	actccccgc	gaatatattta	acacaccgcg	agaagccagct	7260
tattacgctt	ggaaacaacat	atcgcaaaaa	aagattatga	gtttgctgtg	cgtgaattaa	7320
tggtgattct	ggaaaaaaatg	gacgctaatt	ttggaggcgt	tcacgatatt	gaatttgacg	7380
caccgcgtca	gctggcatat	ctaccgcgaa	aattactaat	ttattttgccc	actcgtctcg	7440
ctaattgcga	tacaacacctc	ttttccgacc	ccgaattggc	aatttctgaa	gaagggcgct	7500
taaaatgat	tagcctgcga	cgctggttga	cgtgatattt	tgccctctcc	ccctacgta	7560
acgcagacca	tattctcaat	aaataataa	tcaaccocga	ttccgaaggt	ggctttcaat	7620
tagcaacaga	caactcttct	attgctaaat	tctgtatttt	ttacttacc	gaatccaatg	7680
tcaatgatg	tttagatg	ttatgggcag	ggaatcaaca	actttgtgct	tcaattgtgt	7740
ttgcgttgga	gtcttcaagt	tttattggta	ccgcattctg	gtttcataaa	agagcggtgtg	7800
ttttacagtg	gtttctctaaa	aaactcgccc	aaattgtcaa	tttagatgaa	ttgctcgcaa	7860
atatccttca	tgatgtatat	atgcaactga	gttatgattt	agcaaaaaac	aagcacgatg	7920
ttaaagcttc	attaaacgaa	cttgctcgca	agcatatcct	cacgcgaagg	tggaagaacg	7980
gtcactcttta	caccttaggt	aaaaaggagc	gcaaacctgt	gatgatggta	ctgcttgaac	8040
attttaattc	gggacattcg	atttatcgta	cacattcaac	ttcaatgatt	gctgctcgag	8100
aaaaattcta	tttagctggc	ttaggccatg	agggcggtga	taaaataggt	cgagaaggtg	8160
ttgacgagtt	ctttgaaatc	agtatgaata	atataatgga	tgacgtgttt	tttatccgta	8220
aacagtgcca	aactttccaa	cccgagtggt	tctatatgcc	aagcatggcg	atggatatata	8280
ccagcatttt	ttgtgagcaac	actcggtctg	cccttatcca	actgttagcc	ctgggtcatc	8340
ctgcactac	gcattctgaa	tttattgatt	atgtcatcgt	agaagatgat	tatgtgggca	8400
gtgagaatgt	tttcagcgaa	acccttttac	gttaccocaa	gcttatccca	ccttatgtac	8460
cttctgcaact	cgccccacaa	aaagtggatt	atgtactcag	ggaaaaacct	gaagtgtgta	8520
atatcggtat	tgccgctacc	acaatgaaat	atttttgcta	tttttgcta	acattgcaag	8580
aaatcagaga	taaaagctaaa	gtcaaaaatac	atttttcatt	cgcacttgga	caatcaacag	8640
gcttgacaca	cccttatgtc	aaatgggtta	tggaaagcta	tttagtgta	gatgcccact	8700
cacatcccca	cgcaccttat	cacgattatc	tgccatgatg	cgctgattgc	gatattgctac	8760
taaatccggt	tcttttcggt	aatactaagc	gcataattga	tatggttaca	ttaggtttag	8820
ttgtgtgat	caaaaacgggg	gatgaagtac	atgaacatat	tgagaaggt	cgtgttaaac	8880
gcttaggag	accagaatgg	ctgatagccc	acacacagca	aacatataat	gaatgtgctt	8940
tgcgcttagc	agaaaaacat	caagaacgct	ttgaactcgc	tcgtttacat	atagaaaaa	9000
acggcttaca	aaagcttttt	acaggcgacc	ctcgtccatt	gggcaaaaaa	ctgcttaaga	9060
aaacaaatga	atggaaaggg	aagcacttga	gtaaaaaata	acgggttttt	aaagtataag	9120
tgccgttaat	tttcaaagcg	ttttaaaaac	ctctcaaaaa	tcaaccgcac	ttttatcttt	9180
ataacgatcc	cgcacgctga	cagtttatca	gcctcccgcc	ataaaactcc	gcctttcatg	9240
gcggagattt	tagccaaaac	tggcagaat	taagggttaa	atcacccaaa	ttgcaccaca	9300
aatcaccaa	taccacaaaa	aaa				9323

<210> 7
 <211> 4794
 <212> DNA
 <213> Haemophilus influenzae

<400> 7
 atgaacaaga tatatcgtct caaatccagc aaacgcctga atgctttgggt tgcgtgtgtct 60
 gaattgcacac ggggtgtgtga ccatccaca gaataaggca gtgaaaaacc tgttcgtacg 120
 aaagtacgcc acttggcggtt aaagccactt tccgctatat tgcctatctt gggcatggca 180
 tccattccgc aatctgtttt agcgagcggt ttacagggaa tgagcgtcgt acacggtaca 240
 gcaaccatgc aagtagacgg caataaaacc actatccgta atagcgtcaa tgcctatcat 300
 aattggaaa aattttaacat tgacccaaat gaaatggtag agtttttaca agaaagcagc 360
 aactctgccg ttttcaaccg tttacatct ccaaatatc cccaattaaa agggatttta 420
 gattctaacg gacaagtctt tttaatcaac ccaaatggta tcaacaatagg taaagacgca 480
 attattaca ctaatggctt tactgtctct acgctagaca tttctaacga aaacatcaag 540
 gcgcgttaatt tcacccctga gcaaaccaag gataaagcac tgcgtgaaat cgtgaaatcac 600
 ggttttaata ccgttgggtaa agacggtagc gtaaacctta ttgggtggcaa agtgaataaac 660
 gagggcggtga ttacgctaaa tggcggttag attctttac ttgcagggca aaaaatcacc 720
 atcagcgata taataaaacc aaccatcact tacagcatgt ctgcacctga aaacggaagc 780
 atcaatctgg cgcgatatct tgcacaaagg ggtaacatta atgtccgcgc tgcactatt 840
 cgcaataaag gtaaaccttt tgcgcactct ggaagcaaa ataaaaagtg taacttgtt 900
 ctctctgcc aagaagggtga agcggaattt ggcggtgtta ttctcgctca aaatcagcaa 960
 gccaaagggt gtaagggtgat gattacagcg cataaagtta cattgaaaaa ggggtgcagt 1020
 atcgaccttt cgggttaaga agggggagaa acttatcttg gcggtgacga cgtgtggcaa 1080
 ggtaaaaacg gcaattcaatt agcaaaagaa accactttag aaaaaggctc aacaattaat 1140
 gtgtcaggtg aagaaaaagg tggcgcgct attgattgg gcgattatgc gttaatgcac 1200
 ggcaatatta atgcccaagg taaagatatc gctaaaaact gtggttttgc ggagacgtgc 1260
 gggcattact tatccattga tgaataacga actgttaaaa ctgtctcgca gctactagac 1320
 ccagagaagt tgaactattga agctctctcc gcttctcgcg tcgagctggg tgcgatagg 1380
 aatccccact cggcagaggt gataaaagtg accctaaaaa aaaataaacac ctccctgaca 1440
 acactaacca atacaaccat ttcaaatctt ctgaaaagtg cccacgtggt gaacataacg 1500
 caagggagaa aactaccgt taatagctct atcagtatag aaagaggctc ccacttaatt 1560
 ctccacagtg aaggtcaggg cggctcaaggt gttcagattg ataaagatat tacttctgaa 1620
 ggcggaatt taaccattta ttctggcgga tgggttgatg ttcataaaaa tattacgctt 1680
 ggttagcggt ctttaaacat cacaactaaa gaaggagata tgcctctcga agacaagctt 1740
 ggaaggaaac acttaaacat tacagcccaa gggaccatca cctcaggtaa tagtaacggc 1800
 tttagattta acaacgtctc tctaaacagc ctctggcgaa agctgagctt tactgacagc 1860
 agagaggaca gaggtagaag aactaagggt aatctctcaa acaaatgtga cggaaaggtta 1920
 aactattccg gaactgtaga tatctcaatg aaagcaccca aagtcagctg gttttacaga 1980
 gcaaaaaggac gcacctactg gaacgttaacc actttaaatg ttacctcggt tagtaattt 2040
 aacctctcca ttgacagcac aggaagtggt tcaacaggtc caagcatagc caatgcagaa 2100
 ttaaatggca taacatttaa taaagccact tttaactctg cacaaggctc aacagctaac 2160
 tttagcatca aggcattcaat aatgcccttt aagagttaac ctaactacgc attatttaatt 2220
 gaagattatt cagctctcagg ggggggtagc ctttaattca aacttaacgc ctactcagc 2280
 aacatacaaaa cccctggcgt aattataaaa tctcaaaact ttaattgtctc agggagggtca 2340
 actttaaact tcaagggtga aggttcaaca gaaacgcgct tttcaataga aaatgattta 2400
 aacttaaacg ccaacgggtg caataataca atcagacaag tcgagggtga cgtatcacgc 2460
 gtcaacaaag gtgtcgagcg caaaaaaaac ataactttta aagggggtaa tactaccttc 2520
 ggctctcaaa aagccacaac agaaatcaaa ggcaattgta ccatcaataa aaacactaac 2580
 gctactcttt ttggtgcgaa ttttgccgaa aacaaatcgc ctttaaatat agcaggaaat 2640
 gttatttaata atggcaacct taccactgcc ggtccattta tcaattatgc cggaaattct 2700
 actggtttcaa aaggcgctaa ccttcaagct ataacaattt acacttttaa tgtagccggc 2760
 tcatttgaca acaatggcgc ttcaaacatt tccattgcca gaggaggggc cgtgaaccac 2820
 gatatacaaa accaccagtag cttaaattat accaccaact ctgataccac ttaccgcacc 2880
 attataaaag gcaatatact caacaaatca ggtgatttga atattattga taaaaaaagc 2940
 gcgctgaaa tccaatttgg cggcaatatc caacgaaatc caactttct 3000
 tctgataaag taaattattc caatcagata acaatcaaa gaggcggtga agggggcgct 3060

```

tctgattcaa gtgaggcaga aatgctaac ctaactattc aaaccaaga gttaaaaattg 3120
gcaggagacc taaatatttc aggtctttaa aaagcagaaa ttacagctaa aaatggcagt 3180
gatttaacta ttggcaatgc tagcggttgt aatgctgatg ctaaaaaagt gacttttgac 3240
aagggttaag attcaaaaat ctgcactgac ggtcacaatg taacactaaa tagcgaaagt 3300
aaaaagctta atggtagtag caatgcttgt aatgataaca gcaccggttt aaccattccc 3360
gcaaaagatg taacggtaaa caataacggtt acctcccaca agacaataaa tatctctgcc 3420
gcagcaggaa atgtaacaac caaagaaggc acaactatca atgcaaccac aggcagcgtg 3480
gaagttaact ctcaaaatgg tacaattaaa ggcaacatta cctcgcaaaa tgtaacagt 3540
acagcaacag aaaaatttgt taccacagag aatgctgtca ttaatgcaac cagcggcaca 3600
gtaaacatta gtacaaaac aggggatatt aaaggtggaa ttgaatcaac ttccggtaat 3660
gtaaatatta cagcgagcgg caatacactt aaggtaaagta atactactgg tcaagtgtgta 3720
acagtaacag cggatgcagg agccttgaca actacagcag gctcaacctt tagtgcgaca 3780
acaggcaatg caaatattac aacaaaaaca ggtgatatca acggtaaatg tgaatccagc 3840
tcggctctgt taacacttgt tgcaactgga gcaactctgt ctgtaggtaa tatttcagg 3900
aacactgtta ctattactgc ggatagcggg aaattaacct ccacagtagg ttctacaatt 3960
aatgggacta atagtgtaac cacctcaagc caatcaggcg atattgaaag tacaatttct 4020
ggtaatacag taaatgttac agcaagcact ggtgatttaa ctattggaaa tagtgcaaaa 4080
gttgaagcga aaaaaggagc tgcaacctta actgctgaat caggcaaaat aaccacccaa 4140
acaggctcta cgttaccttc aagcaatggg cagacaactc ttacagccaa gtagagcagt 4200
atcgaggaa acattaatgc tgctaattgt acgttaataa ccacaggcac tttactact 4260
acaggggatt caaagattaa cgcaaccagt ggtaccttaa caatcaatgc aaaaagatgc 4320
aaattagatg gtgctgcatt aggtgaccgc acagtagtaa atgcaactaa cgcaagtggc 4380
tcgtgtaacg tgactgcgaa aacctcaagc agcgtgaata tcaccgggga tttaaacaca 4440
ataaatggct taaatctcat ttccgaaaat gtagaataca ctgtgcgctt aagagggcaag 4500
gaaattgtat tgaatatata ccaaccagggt gtagcaagcg tagaagaggt aatgaaagc 4560
aaacgcgtcc ttgagaaggt aaaaagttaa ctgtgatgaa aaagagaaac actagccaaa 4620
cttgggtgtaa gtgctgtacg tttcgttgag ccaaatatag ccattacggt taatacaca 4680
aacaggttta caaccaaac atcaagtcaa ctgacaattt ctgaaggtaa ggcggtttc 4740
tcaagtgtga atgggcacg agtatgtacc aatgttgctg acgatggaca gcag 4794

```

<210> 8

<211> 4803

<212> DNA

<213> Haemophilus influenzae

<400> 8

```

atgaacaaga tatatcgtct caaattcagc aaacgcctga atgcttttgtg tgcgtgtgtct 60
gaattgcacac ggggttgtga ccatccaca gaataaaagca gtgaaaaaac tggtctgcatc 120
aaagtacgcc acttgcgctt aaagccactt tccgtatatat tgctattctt gggcatggca 180
tccattccgc aatctgtttt agcagcgggt ttacagggaa tgagcgtcgt acacgggtaca 240
gcacacatgc aaagtacagg caataaaaac actactcgta atagcgtcaa tgcatactc 300
aattggaaac aatttaacat tgaccaaaat gaaatggtgc agtttttaca agaagagcag 360
aactctgcgc ttttcaaccg ttgtacatct gaccaaatct ccaattaaa aggtgatttta 420
gattctaacg gacaagtcct tttaatcaac ccaaatggta tcacaatagg taaagagcga 480
attattaaac ctaattggct tactgctctt acgctagaca tttctaacga aaacatcaag 540
gcgcgttaatt tcacccttga gcaaaacaa gataaaagca tcgctgaaat cgtgaatcac 600
gggttaattta cgttggtgaa agacggtagc gtaaacctta ttggtggcaa agtgaataac 660
gagggcggtga ttagcgtaaa tggcggtagt attcttttac ttgcaggga aaaaatcacc 720
atcagcgata taataaatcc aaccatcact tacagcattg ctgcacctga aaacgaagcg 780
atcaactctgg cggatatttt tgccaaaagg ggtaacatta atgtccgcgc tgccactatt 840
cgcaataaag gtaaaacttc tgccgactct gtaagcaaa gataaaagtg taacattgtt 900
ctctctgccca aagaagggtga agcggaaatt ggcgggtgaa tttccgctca aaatcagcaa 960
gccaaaagtg gtaagttgat gattacagggt gataaaagtc acttatcttg cgggtgatga gctgtggcaa 1020
atcgaccttt caggttaaaga agggggagag acctctcttg accctcttag aaaaagctc gcaatttaat 1140
ggtaaaaaatg gatttcaatt agcgaagaaa attgtatggg gcgatattgc attaattaat 1200
gtatcaggca aagaaaaggc cgggcgcgct gctaaaactg gcggttgtt ggaacatca 1260
ggatcatgact tatccattgg tgatgatgtg atgttgtagc ctaaaagtg gttattagac 1320

```

ccagatgatg	tgtccattga	aactcttaca	tctggacgca	ataataccgg	cgaaaaccaa	1380
ggatatacaa	caggagatgg	gactaaagag	tcacctaaag	gtaatatgat	ttctaaacct	1440
acattaacaa	actcaactct	tgagcaaatc	ctaagaagag	tgtcttatgt	taatatcaact	1500
gctcaataata	gaatttatgt	taatagtctc	atcaacttat	ctaattggcg	tttaacacct	1560
cacactaaac	gagatggagt	taaaattaac	ggtgatatta	cctcaaacga	aaatggtaat	1620
ttaacattta	aagcaggctc	ttgggttgat	gttcataaaa	acatcacgct	tggtacgggt	1680
tttttgaata	ttgtcgctgg	ggattctgtg	gcttttgaga	gagaggcgga	taaacgacct	1740
aacgcacaag	atgcctcaat	taccgcacaa	gggacgataa	cgtcaataaa	agatgatata	1800
caatttagat	tcaataatgt	atctattaac	gggacggcca	aggggtttaa	gtttatttga	1860
aatcaaaata	atttcaacta	tcaatttgat	ggcgaaaatt	acatatctgg	aatagtataa	1920
attaaccaaa	ccacgaaaaa	agatgtttaa	tactggaaat	catcaaaaga	ctcttactgg	1980
aatgtttctt	ctcttacttt	gaatacgggt	caaaaattta	cttttataaa	attcgttgat	2040
agcggctcaa	attcccaaga	tttgaggtca	tcacgtagaa	gttttgcagg	cgtacatttt	2100
aacggcatcg	gaggcaaaac	aaacttcaac	atcggagcta	acgcaaaagc	cttattttaa	2160
ttaaaaacaa	acgcgcgtac	agacccaaaa	aaagaattac	ctattacttt	taacgccaac	2220
attacagcta	ccggttaacg	tgatagctct	gtgatgtttg	acatcacagc	caactctaac	2280
tctagagctg	ccggcataaa	catggattca	attacattta	ccggcgggct	tgactttacc	2340
ataaactccc	ataactcgaa	tagtaatgct	tttgaaatca	aaaaagacct	aaactataat	2400
gcaactggct	tgtaattttg	tcttaagcaa	acgaaaagatt	ctttttataa	tgaatacagc	2460
aaacacgcca	ttaactcaag	tcataatcta	accattcttg	gcggaatgtg	caactctagg	2520
ggggaaaaat	caagcagtag	cattacgggc	aatatcaata	tcacaaataa	agcaaatgtt	2580
acattacaag	ctgacaccag	caacagcaac	acaggcttga	agaaaagaac	tcttaactct	2640
ggcaatatat	ctgttgaggg	gaatttaagc	ctaactgggt	caaatgcmaa	cattgtcggc	2700
aatctttcta	tgtcgagaaga	ttccacattt	aaaggagaag	ccagtgcaca	ctctaacatc	2760
acgggcacct	ttaccaacaa	cggtaaccgc	aacattataa	taaaaacaag	agtggtataa	2820
ctccaaaggc	attttattca	taaaagtggt	ttaattatca	ctactaacgc	ctcaggcgact	2880
caaaaaccaa	ttatttaacg	aaataaact	aacgaaaaag	gcgacttaaa	catcaagaat	2940
attaaagccg	acgcgcgaat	ccaaattggc	ggcaatatct	caaaaaaaga	agcgaatctc	3000
acaatttctt	ctgataaagt	aaatattacc	aatcagataa	caatcaaaag	agggcttgaa	3060
ggggggcggt	ctgattcaag	tgaggcgaga	aatgcttaac	taactattca	aaccaaaagat	3120
ttaaaaattg	caggagacct	aaatatttca	ggcttttaata	aagcagaat	tacagctaaa	3180
aatggcagtg	atttaactat	tggcaatgct	agcgttggtg	atgctgatgc	taaaaaagtg	3240
acttttgaca	aggttaaaga	ttcaaaaatc	tgcactgacg	gtcacaatgt	aacactaaat	3300
agcgaagtga	aaacgtctaa	tggtagtatg	aatgctggta	atgataaacg	cacgggttta	3360
accattttccg	caaaaagtgt	aacggtaaac	aaataacgtta	cctcccacaa	gacaataaat	3420
atctctgcgc	cagcaggaaa	tgtaacaacc	aaagaaggca	caactatcaa	tgcaaccaca	3480
ggcagcgttg	aagttaactgc	tcaaaattggt	acaattaaag	gcaacattac	ctcgcaaaat	3540
gtaacagctg	cagcaacaga	aaatcttggt	accacagaga	atgctgtcat	taatgcaacc	3600
agcgcacag	taaacattag	tacaaaaaca	ggggatatta	aaggtggaat	tgaattcaact	3660
tcggtaaatg	taaatattac	agcgagcggc	aaatacactta	aggttaagta	tatcactggt	3720
caagatgtaa	cagtaaacagc	ggatgcagga	gccttgacaa	ctacagcagg	ctcaaccatt	3780
agtgcgacaa	caggcaatgc	aaatattaca	accaaaacag	gtgatatcaa	cggtaaagtt	3840
gaatccagct	ccggctctgt	aaactctggt	gcaactggag	caactcttgc	tgtaggtaat	3900
atttcagagta	acactggttac	tattactgcg	gatagcggtg	aattaacctc	cacagtgaggt	3960
tctcaaatta	atgggactaa	tagtgtaaac	acctcaagcc	aatcaggcga	tattgaaggt	4020
acaattttctg	gtaatacagt	aaatgttaca	gcaagcactg	gtgatttaac	tattggaatt	4080
agtgcaaaaag	ctgaagcgaa	aaatggagct	gcaaccttaa	ctgctgaatc	agggcaaat	4140
accacccaaa	caggctctag	cattacctca	agcaatggtc	agacaaacct	ctgagccaa	4200
gatagcagta	tcgcaggaaa	cattaatgct	gctaatttga	cgtaataaac	ccagcgcaact	4260
ttactactata	caggggattc	aaagattaac	gcaaccagtg	gtaccttaac	aatcaatgca	4320
aaagatgccca	aattgatgtg	tgctgcatca	ggtgaccgca	cagtagttaa	tgcactaac	4380
gcaagtggct	ctggtaacgt	gactgcgaaa	acctcaagca	gcgtgaatat	ccacggggat	4440
ttaaacacaa	aaattgggtt	aaatatcatt	tcggaaaaat	tagaaaacac	gttgcgctta	4500
agaggcaagg	aaattgatgt	gaaatatatc	caaccagggt	tagcaagcgt	agaagaggtg	4560
attgaagcga	aacgcgtctc	tgagaaggta	aaagatttat	ctgatgaaga	aagagaaaca	4620
ctgagcaaac	ctggctgaag	tgctgtacgt	ttcgttgagc	caaaataatg	cattacgggt	4680
aatcacacaaa	acgagtttac	aaccaaacca	tcaagtcaag	tgacaatttc	tgaaggtaag	4740
gcgtgtttct	caagtggtaa	tggcgacaga	gtatgtacca	atgttgtctga	cgatggacag	4800
cag						4803

<210> 9
 <211> 1599
 <212> PRT
 <213> Haemophilus influenzae

<400> 9
 Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu
 1 5 10 15
 Val Ala Val Ser Glu Leu Thr Arg Gly Cys Asp His Ser Thr Glu Lys
 20 25 30
 Gly Ser Glu Lys Pro Val Arg Thr Lys Val Arg His Leu Ala Leu Lys
 35 40 45
 Pro Leu Ser Ala Ile Leu Leu Ser Leu Gly Met Ala Ser Ile Pro Gln
 50 55 60
 Ser Val Leu Ala Ser Gly Leu Gln Gly Met Ser Val Val His Gly Thr
 65 70 75 80
 Ala Thr Met Gln Val Asp Gly Asn Lys Thr Thr Ile Arg Asn Ser Val
 85 90 95
 Asn Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met
 100 105 110
 Glu Gln Phe Leu Gln Glu Ser Ser Ala Val Phe Asn Arg Val
 115 120 125
 Thr Ser Asp Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly
 130 135 140
 Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
 145 150 155 160
 Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
 165 170 175
 Glu Asn Ile Lys Ala Arg Asn Phe Thr Leu Glu Gln Thr Lys Asp Lys
 180 185 190
 Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
 195 200 205
 Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
 210 215 220
 Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
 225 230 235 240
 Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
 245 250 255
 Glu Asn Glu Ala Ile Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn
 260 265 270

Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Lys Gly Lys Leu Ser Ala
 275 280 285
 Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
 290 295 300
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
 305 310 315 320
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
 325 330 335
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
 340 345 350
 Leu Gly Gly Asp Glu Arg Gly Gly Lys Asn Gly Ile Gln Leu Ala
 355 360 365
 Lys Lys Thr Thr Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
 370 375 380
 Glu Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
 385 390 395 400
 Gly Asn Ile Asn Ala Gln Gly Lys Asp Ile Ala Lys Thr Gly Gly Phe
 405 410 415
 Val Glu Thr Ser Gly His Tyr Leu Ser Ile Asp Asp Asn Ala Ile Val
 420 425 430
 Lys Thr Lys Glu Trp Leu Leu Asp Pro Glu Asn Val Thr Ile Glu Ala
 435 440 445
 Pro Ser Ala Ser Arg Val Glu Leu Gly Ala Asp Arg Asn Ser His Ser
 450 455 460
 Ala Glu Val Ile Lys Val Thr Leu Lys Lys Asn Asn Thr Ser Leu Thr
 465 470 475 480
 Thr Leu Thr Asn Thr Thr Ile Ser Asn Leu Leu Lys Ser Ala His Val
 485 490 495
 Val Asn Ile Thr Ala Arg Arg Lys Leu Thr Val Asn Ser Ser Ile Ser
 500 505 510
 Ile Glu Arg Gly Ser His Leu Ile Leu His Ser Glu Gly Gln Gly Gly
 515 520 525
 Gln Gly Val Gln Ile Asp Lys Asp Ile Thr Ser Glu Gly Gly Asn Leu
 530 535 540
 Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn Ile Thr Leu
 545 550 555 560
 Gly Ser Gly Phe Leu Asn Ile Thr Thr Lys Glu Gly Asp Ile Ala Phe
 565 570 575
 Glu Asp Lys Ser Gly Arg Asn Asn Leu Thr Ile Thr Ala Gln Gly Thr
 580 585 590

Ile Thr Ser Gly Asn Ser Asn Gly Phe Arg Phe Asn Asn Val Ser Leu
 595 600 605
 Asn Ser Leu Gly Gly Lys Leu Ser Phe Thr Asp Ser Arg Glu Asp Arg
 610 615 620
 Gly Arg Arg Thr Lys Gly Asn Ile Ser Asn Lys Phe Asp Gly Thr Leu
 625 630 635 640
 Asn Ile Ser Gly Thr Val Asp Ile Ser Met Lys Ala Pro Lys Val Ser
 645 650 655
 Trp Phe Tyr Arg Asp Lys Gly Arg Thr Tyr Trp Asn Val Thr Thr Leu
 660 665 670
 Asn Val Thr Ser Gly Ser Lys Phe Asn Leu Ser Ile Asp Ser Thr Gly
 675 680 685
 Ser Gly Ser Thr Gly Pro Ser Ile Arg Asn Ala Glu Leu Asn Gly Ile
 690 695 700
 Thr Phe Asn Lys Ala Thr Phe Asn Ile Ala Gln Gly Ser Thr Ala Asn
 705 710 715 720
 Phe Ser Ile Lys Ala Ser Ile Met Pro Phe Lys Ser Asn Ala Asn Tyr
 725 730 735
 Ala Leu Phe Asn Glu Asp Ile Ser Val Ser Gly Gly Gly Ser Val Asn
 740 745 750
 Phe Lys Leu Asn Ala Ser Ser Ser Asn Ile Gln Thr Pro Gly Val Ile
 755 760 765
 Ile Lys Ser Gln Asn Phe Asn Val Ser Gly Gly Ser Thr Leu Asn Leu
 770 775 780
 Lys Ala Glu Gly Ser Thr Glu Thr Ala Phe Ser Ile Glu Asn Asp Leu
 785 790 795 800
 Asn Leu Asn Ala Thr Gly Gly Asn Ile Thr Ile Arg Gln Val Glu Gly
 805 810 815
 Thr Asp Ser Arg Val Asn Lys Gly Val Ala Ala Lys Lys Asn Ile Thr
 820 825 830
 Phe Lys Gly Gly Asn Ile Thr Phe Gly Ser Gln Lys Ala Thr Thr Glu
 835 840 845
 Ile Lys Gly Asn Val Thr Ile Asn Lys Asn Thr Asn Ala Thr Leu Arg
 850 855 860
 Gly Ala Asn Phe Ala Glu Asn Lys Ser Pro Leu Asn Ile Ala Gly Asn
 865 870 875 880
 Val Ile Asn Asn Gly Asn Leu Thr Thr Ala Gly Ser Ile Ile Asn Ile
 885 890 895

Ala Gly Asn Leu Thr Val Ser Lys Gly Ala Asn Leu Gln Ala Ile Thr
 900 905 910
 Asn Tyr Thr Phe Asn Val Ala Gly Ser Phe Asp Asn Asn Gly Ala Ser
 915 920 925
 Asn Ile Ser Ile Ala Arg Gly Gly Ala Lys Phe Lys Asp Ile Asn Asn
 930 935 940
 Thr Ser Ser Leu Asn Ile Thr Thr Asn Ser Asp Thr Thr Tyr Arg Thr
 945 950 955 960
 Ile Ile Lys Gly Asn Ile Ser Asn Lys Ser Gly Asp Leu Asn Ile Ile
 965 970 975
 Asp Lys Lys Ser Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser Gln
 980 985 990
 Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Val Asn Ile Thr Asn
 995 1000 1005
 Gln Ile Thr Ile Lys Ala Gly Val Glu Gly Gly Arg Ser Asp Ser Ser
 1010 1015 1020
 Glu Ala Glu Asn Ala Asn Leu Thr Ile Gln Thr Lys Glu Leu Lys Leu
 1025 1030 1035 1040
 Ala Gly Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr Ala
 1045 1050 1055
 Lys Asn Gly Ser Asp Leu Thr Ile Gly Asn Ala Ser Gly Gly Asn Ala
 1060 1065 1070
 Asp Ala Lys Lys Val Thr Phe Asp Lys Val Lys Asp Ser Lys Ile Ser
 1075 1080 1085
 Thr Asp Gly His Asn Val Thr Leu Asn Ser Glu Val Lys Thr Ser Asn
 1090 1095 1100
 Gly Ser Ser Asn Ala Gly Asn Asp Asn Ser Thr Gly Leu Thr Ile Ser
 1105 1110 1115 1120
 Ala Lys Asp Val Thr Val Asn Asn Asn Val Thr Ser His Lys Thr Ile
 1125 1130 1135
 Asn Ile Ser Ala Ala Ala Gly Asn Val Thr Thr Lys Glu Gly Thr Thr
 1140 1145 1150
 Ile Asn Ala Thr Thr Gly Ser Val Glu Val Thr Ala Gln Asn Gly Thr
 1155 1160 1165
 Ile Lys Gly Asn Ile Thr Ser Gln Asn Val Thr Val Thr Ala Thr Glu
 1170 1175 1180
 Asn Leu Val Thr Thr Glu Asn Ala Val Ile Asn Ala Thr Ser Gly Thr
 1185 1190 1195 1200
 Val Asn Ile Ser Thr Lys Thr Gly Asp Ile Lys Gly Gly Ile Glu Ser
 1205 1210 1215

Thr Ser Gly Asn Val Asn Ile Thr Ala Ser Gly Asn Thr Leu Lys Val
 1220 1225 1230
 Ser Asn Ile Thr Gly Gln Asp Val Thr Val Thr Ala Asp Ala Gly Ala
 1235 1240 1245
 Leu Thr Thr Thr Ala Gly Ser Thr Ile Ser Ala Thr Thr Gly Asn Ala
 1250 1255 1260
 Asn Ile Thr Thr Lys Thr Gly Asp Ile Asn Gly Lys Val Glu Ser Ser
 1265 1270 1275 1280
 Ser Gly Ser Val Thr Leu Val Ala Thr Gly Ala Thr Leu Ala Val Gly
 1285 1290 1295
 Asn Ile Ser Gly Asn Thr Val Thr Ile Thr Ala Asp Ser Gly Lys Leu
 1300 1305 1310
 Thr Ser Thr Val Gly Ser Thr Ile Asn Gly Thr Asn Ser Val Thr Thr
 1315 1320 1325
 Ser Ser Gln Ser Gly Asp Ile Glu Gly Thr Ile Ser Gly Asn Thr Val
 1330 1335 1340
 Asn Val Thr Ala Ser Thr Gly Asp Leu Thr Ile Gly Asn Ser Ala Lys
 1345 1350 1355 1360
 Val Glu Ala Lys Asn Gly Ala Ala Thr Leu Thr Ala Glu Ser Gly Lys
 1365 1370 1375
 Leu Thr Thr Gln Thr Gly Ser Ser Ile Thr Ser Ser Asn Gly Gln Thr
 1380 1385 1390
 Thr Leu Thr Ala Lys Asp Ser Ser Ile Ala Gly Asn Ile Asn Ala Ala
 1395 1400 1405
 Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Gly Asp Ser
 1410 1415 1420
 Lys Ile Asn Ala Thr Ser Gly Thr Leu Thr Ile Asn Ala Lys Asp Ala
 1425 1430 1435 1440
 Lys Leu Asp Gly Ala Ala Ser Gly Asp Arg Thr Val Val Asn Ala Thr
 1445 1450 1455
 Asn Ala Ser Gly Ser Gly Asn Val Thr Ala Lys Thr Ser Ser Ser Val
 1460 1465 1470
 Asn Ile Thr Gly Asp Leu Asn Thr Ile Asn Gly Leu Asn Ile Ile Ser
 1475 1480 1485
 Glu Asn Gly Arg Asn Thr Val Arg Leu Arg Gly Lys Glu Ile Asp Val
 1490 1495 1500
 Lys Tyr Ile Gln Pro Gly Val Ala Ser Val Glu Glu Val Ile Glu Ala
 1505 1510 1515 1520

Lys Arg Val Leu Glu Lys Val Lys Asp Leu Ser Asp Glu Glu Arg Glu
 1525 1530 1535

Thr Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Val Glu Pro Asn
 1540 1545 1550

Asn Ala Ile Thr Val Asn Thr Gln Asn Glu Phe Thr Thr Lys Pro Ser
 1555 1560 1565

Ser Gln Val Thr Ile Ser Glu Gly Lys Ala Cys Phe Ser Ser Gly Asn
 1570 1575 1580

Gly Ala Arg Val Cys Thr Asn Val Ala Asp Asp Gly Gln Gln Pro
 1585 1590 1595

<210> 10

<211> 1600

<212> PRT

<213> Haemophilus influenzae

<400> 10

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu
 1 5 10 15

Val Ala Val Ser Glu Leu Thr Arg Gly Cys Asp His Ser Thr Glu Lys
 20 25 30

Gly Ser Glu Lys Pro Val Arg Thr Lys Val Arg His Leu Ala Leu Lys
 35 40 45

Pro Leu Ser Ala Ile Leu Leu Ser Leu Gly Met Ala Ser Ile Pro Gln
 50 55 60

Ser Val Leu Ala Ser Gly Leu Gln Gly Met Ser Val Val His Gly Thr
 65 70 75 80

Ala Thr Met Gln Val Asp Gly Asn Lys Thr Thr Ile Arg Asn Ser Val
 85 90 95

Asn Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met
 100 105 110

Glu Gln Phe Leu Gln Glu Ser Ser Asn Ser Ala Val Phe Asn Arg Val
 115 120 125

Thr Ser Asp Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly
 130 135 140

Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
 145 150 155 160

Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
 165 170 175

Glu Asn Ile Lys Ala Arg Asn Phe Thr Leu Glu Gln Thr Lys Asp Lys
 180 185 190

Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
 195 200 205

Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
 210 215 220

Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
 225 230 235 240

Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
 245 250 255

Glu Asn Glu Ala Ile Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn
 260 265 270

Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Lys Gly Lys Leu Ser Ala
 275 280 285

Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
 290 295 300

Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
 305 310 315 320

Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
 325 330 335

Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
 340 345 350

Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala
 355 360 365

Lys Lys Thr Thr Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
 370 375 380

Glu Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
 385 390 395 400

Gly Asn Ile Asn Ala Gln Gly Ser Asp Ile Ala Lys Thr Gly Gly Phe
 405 410 415

Val Glu Thr Ser Gly His Asp Leu Ser Ile Gly Asp Asp Val Ile Val
 420 425 430

Asp Ala Lys Glu Trp Leu Leu Asp Pro Asp Asp Val Ser Ile Glu Thr
 435 440 445

Leu Thr Ser Gly Arg Asn Asn Thr Gly Glu Asn Gln Gly Tyr Thr Thr
 450 455 460

Gly Asp Gly Thr Lys Glu Ser Pro Lys Gly Asn Ser Ile Ser Lys Pro
 465 470 475 480

Thr Leu Thr Asn Ser Thr Leu Glu Gln Ile Leu Arg Arg Gly Ser Tyr
 485 490 495

Val Asn Ile Thr Ala Asn Asn Arg Ile Tyr Val Asn Ser Ser Ile Asn
 500 505 510

Leu Ser Asn Gly Ser Leu Thr Leu His Thr Lys Arg Asp Gly Val Lys
 515 520 525
 Ile Asn Gly Asp Ile Thr Ser Asn Glu Asn Gly Asn Leu Thr Ile Lys
 530 535 540
 Ala Gly Ser Trp Val Asp Val His Lys Asn Ile Thr Leu Gly Thr Gly
 545 550 555 560
 Phe Leu Asn Ile Val Ala Gly Asp Ser Val Ala Phe Glu Arg Glu Gly
 565 570 575
 Asp Lys Ala Arg Asn Ala Thr Asp Ala Gln Ile Thr Ala Gln Gly Thr
 580 585 590
 Ile Thr Val Asn Lys Asp Asp Lys Gln Phe Arg Phe Asn Asn Val Ser
 595 600 605
 Leu Asn Gly Thr Gly Lys Gly Leu Lys Phe Ile Ala Asn Gln Asn Asn
 610 615 620
 Phe Thr His Lys Phe Asp Gly Glu Ile Asn Ile Ser Gly Ile Val Thr
 625 630 635 640
 Ile Asn Gln Thr Thr Lys Lys Asp Val Lys Tyr Trp Asn Ala Ser Lys
 645 650 655
 Asp Ser Tyr Trp Asn Val Ser Ser Leu Thr Leu Asn Thr Val Gln Lys
 660 665 670
 Phe Thr Phe Ile Lys Phe Val Asp Ser Gly Ser Asn Gly Gln Asp Leu
 675 680 685
 Arg Ser Ser Arg Arg Ser Phe Ala Gly Val His Phe Asn Gly Ile Gly
 690 695 700
 Gly Lys Thr Asn Phe Asn Ile Gly Ala Asn Ala Lys Ala Leu Phe Lys
 705 710 715 720
 Leu Lys Pro Asn Ala Ala Thr Asp Pro Lys Lys Glu Leu Pro Ile Thr
 725 730 735
 Phe Asn Ala Asn Ile Thr Ala Thr Gly Asn Ser Asp Ser Ser Val Met
 740 745 750
 Phe Asp Ile His Ala Asn Leu Thr Ser Arg Ala Ala Gly Ile Asn Met
 755 760 765
 Asp Ser Ile Asn Ile Thr Gly Gly Leu Asp Phe Ser Ile Thr Ser His
 770 775 780
 Asn Arg Asn Ser Asn Ala Phe Glu Ile Lys Lys Asp Leu Thr Ile Asn
 785 790 795 800
 Ala Thr Gly Ser Asn Phe Ser Leu Lys Gln Thr Lys Asp Ser Phe Tyr
 805 810 815

Asn Glu Tyr Ser Lys His Ala Ile Asn Ser Ser His Asn Leu Thr Ile
 820 825 830

Leu Gly Gly Asn Val Thr Leu Gly Gly Glu Asn Ser Ser Ser Ser Ile
 835 840 845

Thr Gly Asn Ile Asn Ile Thr Asn Lys Ala Asn Val Thr Leu Gln Ala
 850 855 860

Asp Thr Ser Asn Ser Asn Thr Gly Leu Lys Lys Arg Thr Leu Thr Leu
 865 870 875 880

Gly Asn Ile Ser Val Glu Gly Asn Leu Ser Leu Thr Gly Ala Asn Ala
 885 890 895

Asn Ile Val Gly Asn Leu Ser Ile Ala Glu Asp Ser Thr Phe Lys Gly
 900 905 910

Glu Ala Ser Asp Asn Leu Asn Ile Thr Gly Thr Phe Thr Asn Asn Gly
 915 920 925

Thr Ala Asn Ile Asn Ile Lys Gly Val Val Lys Leu Gly Asp Ile Asn
 930 935 940

Asn Lys Gly Gly Leu Asn Ile Thr Thr Asn Ala Ser Gly Thr Gln Lys
 945 950 955 960

Thr Ile Ile Asn Gly Asn Ile Thr Asn Glu Lys Gly Asp Leu Asn Ile
 965 970 975

Lys Asn Ile Lys Ala Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser
 980 985 990

Gln Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Val Asn Ile Thr
 995 1000 1005

Asn Gln Ile Thr Ile Lys Ala Gly Val Glu Gly Gly Arg Ser Asp Ser
 1010 1015 1020

Ser Glu Ala Glu Asn Ala Asn Leu Thr Ile Gln Thr Lys Glu Leu Lys
 1025 1030 1035 1040

Leu Ala Gly Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr
 1045 1050 1055

Ala Lys Asn Gly Ser Asp Leu Thr Ile Gly Asn Ala Ser Gly Gly Asn
 1060 1065 1070

Ala Asp Ala Lys Lys Val Thr Phe Asp Lys Val Lys Asp Ser Lys Ile
 1075 1080 1085

Ser Thr Asp Gly His Asn Val Thr Leu Asn Ser Glu Val Lys Thr Ser
 1090 1095 1100

Asn Gly Ser Ser Asn Ala Gly Asn Asp Asn Ser Thr Gly Leu Thr Ile
 1105 1110 1115 1120

Ser Ala Lys Asp Val Thr Val Asn Asn Asn Val Thr Ser His Lys Thr
 1125 1130 1135

Ile Asn Ile Ser Ala Ala Ala Gly Asn Val Thr Thr Lys Glu Gly Thr
 1140 1145 1150
 Thr Ile Asn Ala Thr Thr Gly Ser Val Glu Val Thr Ala Gln Asn Gly
 1155 1160 1165
 Thr Ile Lys Gly Asn Ile Thr Ser Gln Asn Val Thr Val Thr Ala Thr
 1170 1175 1180
 Glu Asn Leu Val Thr Thr Glu Asn Ala Val Ile Asn Ala Thr Ser Gly
 1185 1190 1195 1200
 Thr Val Asn Ile Ser Thr Lys Thr Gly Asp Ile Lys Gly Gly Ile Glu
 1205 1210 1215
 Ser Thr Ser Gly Asn Val Asn Ile Thr Ala Ser Gly Asn Thr Leu Lys
 1220 1225 1230
 Val Ser Asn Ile Thr Gly Gln Asp Val Thr Val Thr Ala Asp Ala Gly
 1235 1240 1245
 Ala Leu Thr Thr Thr Ala Gly Ser Thr Ile Ser Ala Thr Thr Gly Asn
 1250 1255 1260
 Ala Asn Ile Thr Thr Lys Thr Gly Asp Ile Asn Gly Lys Val Glu Ser
 1265 1270 1275 1280
 Ser Ser Gly Ser Val Thr Leu Val Ala Thr Gly Ala Thr Leu Ala Val
 1285 1290 1295
 Gly Asn Ile Ser Gly Asn Thr Val Thr Ile Thr Ala Asp Ser Gly Lys
 1300 1305 1310
 Leu Thr Ser Thr Val Gly Ser Thr Ile Asn Gly Thr Asn Ser Val Thr
 1315 1320 1325
 Thr Ser Ser Gln Ser Gly Asp Ile Glu Gly Thr Ile Ser Gly Asn Thr
 1330 1335 1340
 Val Asn Val Thr Ala Ser Thr Gly Asp Leu Thr Ile Gly Asn Ser Ala
 1345 1350 1355 1360
 Lys Val Glu Ala Lys Asn Gly Ala Ala Thr Leu Thr Ala Glu Ser Gly
 1365 1370 1375
 Lys Leu Thr Thr Gln Thr Gly Ser Ser Ile Thr Ser Ser Asn Gly Gln
 1380 1385 1390
 Thr Thr Leu Thr Ala Lys Asp Ser Ser Ile Ala Gly Asn Ile Asn Ala
 1395 1400 1405
 Ala Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Thr Gly Asp
 1410 1415 1420
 Ser Lys Ile Asn Ala Thr Ser Gly Thr Leu Thr Ile Asn Ala Lys Asp
 1425 1430 1435 1440

Ala Lys Leu Asp Gly Ala Ala Ser Gly Asp Arg Thr Val Val Asn Ala
 1445 1450 1455

Thr Asn Ala Ser Gly Ser Gly Asn Val Thr Ala Lys Thr Ser Ser Ser
 1460 1465 1470

Val Asn Ile Thr Gly Asp Leu Asn Thr Ile Asn Gly Leu Asn Ile Ile
 1475 1480 1485

Ser Glu Asn Gly Arg Asn Thr Val Arg Leu Arg Gly Lys Glu Ile Asp
 1490 1495 1500

Val Lys Tyr Ile Gln Pro Gly Val Ala Ser Val Glu Glu Val Ile Glu
 1505 1510 1515 1520

Ala Lys Arg Val Leu Glu Lys Val Lys Asp Leu Ser Asp Glu Glu Arg
 1525 1530 1535

Glu Thr Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Val Glu Pro
 1540 1545 1550

Asn Asn Ala Ile Thr Val Asn Thr Gln Asn Glu Phe Thr Thr Lys Pro
 1555 1560 1565

Ser Ser Gln Val Thr Ile Ser Glu Gly Lys Ala Cys Phe Ser Ser Gly
 1570 1575 1580

Asn Gly Ala Arg Val Cys Thr Asn Val Ala Asp Asp Gly Gln Gln Pro
 1585 1590 1595 1600

<210> 11
 <211> 29
 <212> PRT
 <213> Haemophilus influenzae

<400> 11
 Val Asp Glu Val Ile Glu Ala Lys Arg Ile Leu Glu Lys Val Lys Asp
 1 5 10 15

Leu Ser Asp Glu Glu Arg Glu Ala Leu Ala Lys Leu Gly
 20 25